

3 September 1981

Where now for the British Association?

The British Association, meeting at York this week, has several causes for celebration, the chief of which is that it has lasted for 150 years. Since the Second World War, the association's survival has often been in doubt. Now, mercifully, there seems to be a breathing space. With the help of a modest subvention from the British government (through the Royal Society), a membership which, while stagnant, is at least not falling, and a palpable enthusiasm within the young people's section, the British Association has less need to worry about its immediate future than for many years. But presumably for these reasons, the association should use its birthday party this week to decide where it goes from here and how.

As on other birthdays, the contemplation of the past may be worthwhile — and this week's meeting has been provided with several such opportunities. The retrospective symposium on Wednesday, extracts from which appear on pages 13–32, are a vivid reminder of the distinction and of the importance of the association's origins. Its roots are firmly planted in the characteristically English Industrial Revolution. Its legitimate heroes are the Faradays and Daltons of that world. At the beginning, the association was a forum for the incorrigible (and incorrect) belief of the early Victorians that an understanding of science would enable them to come to grips with an exciting world still largely unexplored. It is sometimes said that if in 1831 the Royal Society had been more active, the British Association would not have been necessary, but that is a mistake. The essence of the association, then as now, is that it is not an academy in the ordinary sense but rather is open to all comers. So it should remain.

The contemplation of these distinctive origins should, however, not blind those in York this week to the bad patches there have been. Although the crises afflicting the association have usually been presented as financial crises — too little money to meet obvious needs — financial problems have been symptoms, not causes. The truth is that in its blackest days the association has lacked a sense of what it should be doing and has as a result lost opportunities for doing good works and for giving its membership a sense of belonging to a socially exciting institution. The association has since the Second World War hardly ever taken the lead in anything that matters — the teaching of science in schools, problems of nuclear weapons or of arms control, problems of environment and pollution, the present issues surrounding the commercial exploitation of genetic manipulation. That it has ceased to be a scientific society in the ordinary sense is understandable — professional societies are too professional — but the association has not often been used as a means by which people at large can be told of those issues which concern professionals but which can be resolved only by involving a wider audience. This week, the association's justifiable self-congratulation should be matched by an attempt to understand why its recent performance has been so second rate. Several remediable failings stand out.

For one thing, the association is badly organized and should be reformed. For the past 30 years, a succession of part-time secretaries have complained that the annual meeting would have

been interesting if power over the meeting programme had not been centred in the hands of people called the section recorders, those volunteer academics who can cajole others into turning up at various provincial cities (with no expenses paid) to give talks to unknown and unpredictable audiences. It is not surprising that most annual meetings are a ragbag, yet the recorders provide virtually the only intellectual continuity in the association's structure. Without the recorders, the association's most public activity would not exist. The failing is not theirs but that of the top management, which has consistently been so preoccupied with money and procedure that it has failed to come to grips with the intellectual questions that should concern the association. But if the management cannot lead the membership, how can the association itself give a lead to a wider audience?

A few simple but radical reforms are necessary. The association should no longer limp on with a chief executive officer (called the secretary) who has made his mark in some other office and who can be counted on not to ask for too much salary because he has pre-existing pension arrangements. The need is for people who are younger and fully committed both to the process of defining a role for the association and then to its implementation. Second, the association must learn to live with democracy. Not being an academy, a body that legitimately perpetuates itself by electing to its membership people chosen privately, the association cannot continue to deny the membership at large the choice of who should manage their association. For the past ten years there has been a kind of electoral system for members of the governing council, but the council itself has very little say in the choice of, for example, the president (who lasts for a year) and of the executive committee consisting of general secretaries, who can (and sometimes almost do) almost go on for ever. If this were not so patently ridiculous, it would be outrageous. The constant fear is that democracy might allow the association to get into the "wrong" hands. But if the members of the association were able to decide for themselves how and by whom it should be managed, the risk that it might shoot off in some foolish direction would be more palatable than the near-certainty, under present arrangements, that the association will collapse because it cannot decide what next to do.

But what should that be? The association's perennial problem has become that of finding a cause to prosecute. The annual meeting happens (and should continue) because the membership wants it. Plans for separate symposia or studies of particular topics — the social consequences of microprocessors or radioactive waste disposal — have often been successful but have been too sporadic to suggest that the association is a serious source of information and comment on problems of that kind problems that perplex the community as a whole. The association likes to think that it helps inform the general public of the importance of science in the modern world, but it has let its regular publications lapse and so has fallen back on word of mouth and the young people's section. And, so far at least, the association has not shown a flicker of interest in the development of the past few years and weeks that may most directly threaten the integrity of scientific enterprise in Britain — the decision on



the pattern of higher education now being enforced.

This week's retrospective view of what the association has done may in this sense help. The record shows that it was not, even in its heroic days, a strictly academic body but one that sought to build bridges between the scientific community and other groups of interested people. Who would say that the need for that service has diminished in the past 150 years? But what has changed is the nature of society. Many of the dreams of the 1830s have come true — people are more healthy, more wealthy and even wiser. Questions such as the validity of evolutionary theories, the ethical problems thrown up by modern biology and the economic future of technological societies concern people of all backgrounds, not just scientists. No single body could resolve all those conundrums. But the British Association is well placed to promote the intelligent discussion of them. It is also better qualified for that task than for its present attempts at popularization at which it is not especially skilled. Is it too much to hope that this week's birthday party will give a substantial proportion of the membership the resolve to set off in this new (but also old) direction?

Mr Watt's fight

One of the colourful characters of the new United States Administration and the most eccentric is Mr James Watt, the Secretary of the Interior. For most of the past six months, Mr Watt has been in one scrape after another and has succeeded in elevating the Department of the Interior from its traditional obscurity to a position of one of the most controversial of government departments. For Mr Watt is a zealot, one who appears to believe that the great resources of the United States should be more fully harnessed to the cause of economic growth, even if the consequences are that this or that patch of wilderness or national park will be sacrificed. The outrage of the environmentalists is predictable. But Mr Watt may now be surprised to find himself in trouble with Congress, not so much for his policies (which nevertheless are not much liked) as for his manner (for his certainty that he is right has offended even some of his potential friends).

Mr Watt's problem is familiar. He is by all accounts a man of high principle (and a deeply religious conviction). He takes the view that mankind has a duty to ensure its own survival, to which end the efficient exploitation of natural resources is essential. High principles are not, of course, unwelcome in politics — indeed, some would say that they are all too rare. What Mr Watt appears not to have appreciated or to have had time to learn (for he had no previous experience of practical politics) is that there are many occasions on which equally high-minded people differ radically in their opinions about courses of action that should be followed. Mr Watt's failure is especially tragic, apart from the trouble he is likely to bring down on himself, because much of what he seeks to do is sensible.

The latest row with which Mr Watt has been caught up concerns a proposal to license drilling for oil and gas in a wilderness area in Montana. The issue is simple. The wilderness area, which spans the Great Divide, is federal land. The statutes which set it aside as wilderness can be amended only by Congress. The chances of finding oil and gas are said to be good, at least on paper, but so far not a single exploratory borehole has been drilled. Mr Watt appears to favour letting a few licences to drill in this remote area of the United States, thus confirming his critics in their view that he is the kind of man who would allow the Lincoln Memorial to be broken up if there were no marble for some other building project.

Mr Watt is right to worry about the supply of oil and natural gas. That is his job. He would also be on firm ground if he were to argue that circumstances are foreseeable in which shortages of energy are so great that environmental considerations of all kinds would have to be set aside. His mistake, on the most generous reading of his actions since taking office, is that he supposes that

his assessment of the urgency of the need takes precedence over that of other people. Although, in the days of President Lyndon Johnson's Great Society, wilderness areas were dedicated and national parks extended without nearly enough thought for the consequences, they were popular accretions; people have grown fond of them. And United States voters will not give them up without a good reason. Browbeating from Mr Watt is almost certain to be counter-productive.

Much will depend in the weeks ahead on whether Mr Watt seeks to win his case or, less prudently, to make his point that conservation must now take a back seat. In the first and prudent case, he will postpone the issue of whether people should begin immediately to drill for oil and gas in Montana, and will instead try to hammer out a rational policy on conservation. When should the need for natural resources take precedence over the conservation of the native or artificial state of some valued tract of land? What conditions must be attached to such projects to strike a proper balance between economic efficiency and natural beauty, the preservation of habitats and of species? By what means can some kind of economic value be put on a tract of wilderness? And how much land can even the United States afford to put in quarantine for ever? These questions were scandalously glossed over in the past two decades, when any tract of federal land not obviously an eyesore was liable to become a conservation area overnight. But, as things are, and however unfair it may seem, Mr Watt will not be able to get his way until he sets out to answer them patiently and persuasively. To do that is the prudent course. The imprudent course, on which he seems bent, is to behave as if those who oppose him have no case. For Mr Watt, the consequence of that will be such a fight with Congress that the Administration will find him a political liability — and will find another Secretary of the Interior. And then there will be no oil and gas from Montana.

Isidore Nabi, RIP

There has been great confusion in the scientific literature because of a jape that began at the University of Chicago some years ago. A non-existent scientist, Dr Isidore Nabi (whose first name is sometimes spelled Isadore), was blessed with a biography in *American Men and Women of Science* by a group of scientists including Professor Leigh Van Valen (still at the University of Chicago), Dr Richard C. Lewontin (now a professor at Harvard University) and Dr Richard Lester (now at the Harvard School of Public Health). Although, no doubt, the editors of *American Men and Women of Science* will be offended to discover that they have been duped, the creation of Nabi from thin air may be thought a harmless joke.

Unfortunately the joke has gone too far. Apparently Nabi's three creators have been in the habit of using his fake existence as a means of concealing their own identity. Earlier this year, for example, a letter supposed to be from Nabi was published in *Nature* (290, 183; 1981) making an otherwise plausible point about the controversy over the Natural History Museum. Nabi's name has also turned up elsewhere, even as the author of articles in the journal called *Science and Nature*. The objection to this use of Nabi's fictional identity as a pseudonym in the scientific literature is twofold. First, it is a deception. Second, it allows people with known opinions on important controversial matters to give a false impression that their opinions are more weighty than truth would allow.

So somehow Nabi has to be banished from the scientific literature. What began as a good joke has become an impediment to sensible discussion. But if Nabi's three creators insist on using his name as a pseudonym, what can simple mortals do? The answer is quite simple — let others than those in the know use Nabi's name frequently, especially when making points conflicting with those who have so far used the pseudonym. It should not be long before they find it necessary to invent another or, better still, to use their own names.

Congress faces decision on CFC

New ozone data from NASA

Washington

When Congress reconvenes next week one of the many contentious issues it will face is the future of US controls on the production and use of chlorofluorocarbons (CFCs). The debate has been given a new twist by two factors: data that seem to provide the first empirical evidence of depletion of the Earth's ozone layer, which might be attributed to CFCs; and an intensified campaign by both manufacturers and industrial users to head off any further regulation.

Having banned the use of CFCs in most aerosol sprays four years ago, the US Environmental Protection Agency (EPA) announced last October that it intended to broaden its restrictions to cover non-aerosol uses such as refrigerants, foaming agents and solvents. The move came after two reports from the National Academy of Sciences appeared to confirm earlier warnings that depletion of the ozone layer by CFCs could cause problems ranging from an increase in human skin cancer due to the extra ultraviolet radiation that the Earth's atmosphere would let through, to unpredictable effects on agricultural yields as a result of climatic changes.

Although EPA still has an official deadline of December 1981 for implementing the new restrictions, which in their present form would ultimately limit production of CFCs in the United States to 30 per cent of the present levels, such moves seem to be in abeyance as part of the Reagan Administration's general effort to reduce the burden of environmental regulation on the private sector.

But the CFC producers want more. At their instigation, draft bills have been introduced into both the Senate and the House of Representatives which would restrain EPA from imposing further controls on the production and use of CFCs until there was "clear scientific evidence" that they pose a threat to humans and the environment.

Until recently, manufacturers had made considerable play on the fact that the predicted effect of CFCs on the ozone layer — which a National Academy of Sciences report two years ago said would be depleted by about 16 per cent if present trends were to continue — have been based primarily on scientific models with little empirical verification. They criticized EPA for creating the impression that the CFC ozone theory is "fact" rather than a "theory" whose calculations "are very much in question". The situation is now changing

following evidence produced by scientists with the National Aeronautics and Space Administration (NASA) of what appears to be a steady depletion of the ozone layer 40 kilometres above the surface of the Earth at a rate compatible with some of the theoretical predictions.

The NASA data are based on readings made from two experimental meteorological satellites — Nimbus Four and Nimbus Seven. Both use a "total ozone mapping spectrometer" which can detect the ozone concentration at different heights above the Earth's surface.

According to Dr Donald Heath, a staff scientist with NASA who has been responsible for designing the ozone experiments, analysis of the measurements over the period 1970–79, once seasonal variations and other factors have been eliminated,

reveal two principal components. The first can be related directly to the 11-year solar cycle; but in addition to this, says Dr Heath, there appears to be a steady decrease of about 0.5 per cent a year at a height of 40 kilometres.

Although Dr Heath says that CFCs "represent the most likely explanation", he admits that other hypotheses could explain his data, such as the effect of different components of the solar spectrum: "the important thing is that apparently the effect [on ozone depletion] predicted by the theory is being confirmed, and a change in the ozone concentration does seem to be taking place".

Industry representatives remain sceptical, pointing out that the NASA data have not yet been subject to peer review or published. They emphasize possible weak-

Britain losing out on information technology

The private telecommunications industry in the United Kingdom appears to be making another strong play for a slice of the British telecommunications cake — this time in the strange guise of a report prepared for the National Enterprise Board (NEB), made public last week.

The report was written by staff of the telecom and computer division of the management and technical consultants PA International, and it warns that the deficit in the British balance of payments in the telecom industry will rise to £1,000 million a year in 1990 — unless the government takes a number of major initiatives.

In prime place among them is the proposal that the Department of Industry should "clarify and implement the liberalization of British telecommunications promptly". This means to make prompt use of the act which recently established the company British Telecom (the Post Office telecommunications division under a new guise). The act withdraws the previous Post Office monopoly in telecommunications — but vests discretion granting project licences in the Secretary of State for Industry, free marketer Sir Keith Joseph. Private industry, backed by the PA report, wants licences now, and is also concerned that under the present act liberalization of the market could be reversed by a Secretary of State who opposed Sir Keith's views (as would probably follow, for example, if a Labour government were installed at the next general election, due by 1984).

Besides liberalization, the NEB-commissioned report calls for British Telecom to be allowed to raise capital on the private market; for government departments which are major spenders or potential major spenders in the information industry to coordinate procurement policies, with a view to stimulating the proper growth and international orientation of the industry; for the

Department of Industry and Ministry of Defence to help UK companies win international bids by financial backing and guarantees for risk-sharing consortia; and for government departments to stimulate technology transfer from the government sector to the commercial sector.

PA recognizes that such wide-sweeping proposals may fall on deaf ears, but hopes they may stimulate some thinking. The trouble is that the proposals smack of a central technology policy, and Britain has no appropriate minister to champion or implement it. The Prime Minister is also on record as saying that she sees no need for such centralization (*Nature* 281,249; 1979).

A different obstacle — government opposition to new spending — may also apply to another group of proposals in the PA report, which advocates a number of special government-led projects in telecommunications. One would be a large-scale demonstration "electronic office"; another an electronic business communications system for the City of London. Existing conflicting commercial interests in these areas make a government "pump-priming" role essential.

Britain must also make a greater effort to find a place in the international market, says the report. British suppliers, with a few exceptions, are heavily dependent on the UK market, and particularly on government procurement. "Moreover government procurement and research and development funding is not being directed as effectively in the United Kingdom as it is elsewhere", it says — casting an envious eye at France, Germany, Japan and the United States. Nevertheless, says PA, there are signs that the government — and British Telecom — is waking up to the fact.

Robert Walgate

● "A Strategy for Information Technology" is available from PACTEL, Rochester House, 33 Greycoat Street, London SW1P 2QF

nesses in the data resulting from the calibration techniques used and alternative hypotheses that do not necessarily implicate CFCs.

However, the NASA data are still likely to be prominent in future debates on CFC controls. The proposed legislation would amend the Clean Air Act of 1970, due for renewal by Congress later this month, to shift the focus of EPA's actions from regulation to research. It would require EPA "to use statistical analysis of actual ozone measurements as an early warning system to guard against excessive ozone depletion", according to the Alliance for Responsible CFC Policy, the industry lobby which has been largely responsible for the proposed legislation.

Environmental groups have expressed strong opposition to the proposed legislation. Giving evidence to an oversight committee of the Senate Environment and Public Works, an attorney with the Natural Resources Defense Council claimed that "legislation creating narrowly-defined factual conditions for the imposition of regulation in any form to protect the public against a serious environmental risk would be bad policy and a dangerous precedent".

The environmentalist group also criticized a provision in the Senate version of the bill which would restrain EPA from taking any further action without an international consensus that more regulation was necessary. In the past the US government has been the first to act on restricting CFCs; Norway, Sweden and Canada have followed with complete bans on aerosols, and the members of the European Economic Community agreed last year to reduce overall production.

National Academy of Sciences officials are no happier with the suggestion that they might be asked to judge whether or not there was sufficient evidence of a link between CFCs, ozone depletion and threats to human health to require additional regulation. They argue that their function should be restricted to evaluating the adequacy of the scientific data.

Industry remains confident both that it can hold the Administration back from any further controls in the short term and that it can persuade Congress to legislate a new approach to the control of CFCs which would survive into a new administration with potentially different values.

EPA has not announced any change in direction from the proposals published last year — but it is clearly sympathetic to the industry's point of view. At her confirmation hearings before the Senate, the new EPA administrator, Mrs Ann Gorsuch, said she understood that the theory of the stratospheric ozone layer was "highly controversial", and that "apparently there is a need for additional scientific data before the international scientific community would be willing to accept the ozone layer depletion theory as the basis for additional government action".

David Dickson

Soviet natural resources

Try harder comrades

The Soviet Union has launched yet another drive against waste of energy and raw materials. A resolution of the Central Committee of the Communist Party of the Soviet Union and the Council of Ministers of the USSR published recently calls for a radical improvement in the savings of such materials in all branches of production and an increase in the role of scientists in effecting these economies.

The new resolution, while praising the work of "leading collectives", admits that there is considerable waste. The consumption of raw materials and energy per unit of national income, it says, is higher than the "best world indices". Also, inefficient mining and extraction processes mean that large quantities of coal, oil and ore remain underground.

To cure these ills, the resolution urges a stringent economy regime, more modern management methods and a mass publicity campaign. The State Committee for Science and Technology, the Academy of Sciences and the various production ministries are to undertake research aimed at reducing the specific energy and resource consumption of industry and maximizing the use of recycled resources.

The ministries involved in education are to improve the level of teaching of economic disciplines, while Party courses and organizations such as the Komsomols are to study the problems from a political theoretic standpoint. The "Znanie" ("Knowledge") society, which coordinates and supports the work of science clubs and societies throughout the Soviet Union, is to prepare material demonstrating the correct use of resources in the economic strategy of the Party. The press, cinema, television, radio and the "Plakat" poster press have all been mobilized in the efficiency drive, and a complex scheme of sanctions for energy wasters and financial incentives for the successful energy savers is being worked out.

Special attention continues to be paid to the petrochemical sector. Earlier this year, the petrochemical industry was hived off from the chemical industry into a ministry of its own. In addition, it has now been decided to establish a special State Committee for the Supply of Oil Products.

At the same time, a crash programme for power station construction has been launched, concentrating on the nuclear sector. At a two-day meeting in mid-July on nuclear power, Dr Anatolii Aleksandrov, president of the Academy of Sciences stressed the need for further automation in the operation of nuclear power stations, and the importance of speeding up the construction of fast-neutron reactors. About 25 million kW of nuclear generating capacity is to be commissioned in the next five years, three times more than during the previous five years.

Vera Rich

Heat on BNFL

The depressed demand for electricity in the United Kingdom and the increasing public distrust of things nuclear are the main worries for the future of state-owned British Nuclear Fuels Limited (BNFL), reflected in the company's tenth annual report published last week. The report reveals a £3 million fall in profits, to £12.9 million, in spite of an increase in sales.

The strength of sterling, together with high interest rates and the inevitable effects of inflation have combined to reduce BNFL's international competitiveness. Defence cuts, and the government's rigorous policy of "cash limits" on BNFL's other main customer, the Central Electricity Generating Board, have reduced the need for enriched uranium and the closure of the outdated diffusion enrichment plant at Capenhurst has been brought forward to the end of 1982.

BNFL chairman Sir John Hill said when presenting the report that further business expansion for BNFL "is tied inextricably to the future of nuclear power as a major energy source and in particular to the restoration of growth of electricity demand in the United Kingdom".

This commitment is underlined by Sir John's efforts to counter the flow of propaganda against the building of new nuclear power stations — "Despite persistent attack from anti-nuclear factions, the case for nuclear power with its inherent safety and economic advantage, remains intact".

Charles Wenz

French nuclear energy

Summer strife

French energy normally flags in August, when the nation takes its traditional holidays; but not so this year. The Mitterrand government is flinging itself towards its first full parliament, wishing to get a number of issues quickly under its belt before there can be too much shouting. But with one of them — nuclear power — it is already having some trouble.

The French environmentalists got short shrift under Giscard d'Estaing. Mitterrand, a socialist, not beholden to big business, appeared to offer them more — he promised a halt to nuclear construction, a national debate and even a referendum. They voted for him.

Now things look a little different: the parliamentary debate on nuclear power will be in October, and may last less than a week. There is no longer talk of a referendum; nuclear construction continues at many sites; and, as if to underline the situation, Mitterrand's

Minister of State for Research and Technology, Jean-Pierre Chevenement, said recently that nuclear power was necessary to France, but that "democracy" was worth a delay of a few months.

So August has seen increasing tension over energy issues — only heightened by the arrival in Cherbourg of a British freighter from Japan, carrying Japanese nuclear waste for reprocessing at Cap de la Hague in Normandy. There was a demonstration against which the police used tear gas. And on the other side, the broadly Communist-led union, the CGT, protested to the Prime Minister against the pause that had been imposed by the government at certain less-advanced reactor construction sites, such as Cattenom on the Moselle near Luxembourg, while at other sites workers and local contractors blocked roads to protest against the pause.

So it seems that energy politics in a France that at least pays lip service to democracy is going to be just as difficult as it has proved in other western countries, and it remains to be seen whether the government can handle the situation in any other than the effectually dirigiste manner adopted by the previous administration.

Much must wait for the debate and its aftermath, but for the moment the government has been shaken into taking one new step — the setting up of a permanent scientific commission to enquire into the operation of the Cap de la Hague reprocessing plant, which has suffered a number of leaks and accidents and has never functioned efficiently.

Prime Minister Pierre Mauroy appeared to have been forced into this position by a broadly-based group which opposes the extension of Cap de la Hague to treat foreign waste. Shortly after the landing of the Japanese cargo, the group was led to believe by a member of Mauroy's cabinet that the extension of the plant and the treatment of foreign waste were inevitable. They came away saying they were deceived and disheartened and that the government's position was "very clear and very grave". Almost immediately Mauroy gave them a personal audience, announced the setting up of the commission and indicated that no foreign fuel would be reprocessed until after the debate.

Where this leaves energy policy in France is, however, unclear. The socialist party appears to be split on the issue. Chevenement is the most outspoken pro-nuclear spokesman. Other ministers are more equivocal, notably Minister of Finance Jacques Delors, who says he does not want France to become the waste-dump of Europe. The debate may only further muddy these waters, for elements of the socialist party — represented by national secretary Paul Quilès — believe that it should not lead to decisions. Rather, Quilès says, it should help to define a democratic framework within which the final decision can be taken.

Robert Walgate

Indian technology Home and away

Bangalore and Lucknow

Although India has a well defined national policy on science, there is still no such luxury for technology. An attempt by the Janata government that ruled India between 1977 and 1979 to formulate a technology policy failed because the draft was never approved by the cabinet. Now the present government led by Mrs Gandhi finds the draft Janata policy "lacking in emphasis on self-reliance", and efforts are now being made to give India a "comprehensive self-reliant" policy on technology. A "special group" has been set up by the Science Advisory Committee of the cabinet to come up with a new draft.

Professor M.G.K. Menon, Secretary of the Science and Technology Department, says the new policy will "reflect the government's stress on self-reliance". Some of the main objectives that will ultimately form part of the policy statement have already been spelt out by the working group on science and technology for India's sixth five-year plan. One new feature is likely to be the idea of maintaining a register of all foreign collaborations. The aim is to ensure that prime contractors for technology policy are Indian and that there is a firm commitment to involve Indian research and development activities whenever foreign technological know-how is being imported.

Professor Menon observed that "the need for formulation and adaptation of a new national policy has been felt for quite some time", and said that state-level science and technology departments would be set up to make sure that the policies formulated at the centre were properly implemented.

Radhakrishna Rao

● India is to set up an agency specifically to regulate the transfer of technology to other developing countries. India already has 207 joint ventures with Third World countries under way abroad — 115 already in operation and 92 in various stages of planning and construction.

India's involvement in these projects is mainly through the export of capital equipment and technology, although there is also some transfer of cash. Most of the effort goes towards light engineering and textiles, but there are also projects in oil fractionation, the paper industry and chemicals and pharmaceuticals.

The proposed new agency is expected to coordinate the efforts of public sector undertakings, such as Engineering Projects (India), Bharat Heavy Electricals and Electronics Trade and Development Corporation. It is to be part of the Department of Science and Technology but will have links with the External Affairs Ministry to help promote cooperation between India and Third World countries.

Zaka Imam

US synthetic fuels Going for broke

Washington

Disregarding several of his top-level economic advisers, President Reagan has approved federal loan guarantees totalling \$3,500 million to support investments by private companies in three plants being built to produce synthetic fuel (synfuel) from coal and oil shale.

The three loan guarantees are the first to have been offered under legislation passed last year by Congress providing up to \$20,000 million a year to support synfuel projects. They had been strongly opposed by Administration economists such as Mr David Stockman, director of the Office of Management and Budget (OMB), on the grounds that the loans represented an unwarranted subsidy for commercial companies.

The synfuels programme was set up by the Carter Administration to accelerate efforts to produce both liquid fuels and natural gas from domestic resources of coal, oil shale, tar sands and heavy oil. The centrepiece of these efforts is to be a Synthetic Fuels Corporation, formally established by last year's legislation, but still to come into effective operation.

In July the Reagan Administration outlined a new energy policy, which seemed to reinforce its proposed withdrawal from issuing large federal subsidies to energy projects, particularly those in the development stage.

The Administration declared that the Department of Energy "will continue to support and fund long-term, high-risk research and development projects which industry would not be in a position to finance", but added that the development pace for the US synthetic fields industry "will be determined by private investors".

Among projects promised loans in the final days of the Carter Administration and which it was argued would collapse without government support, is a coal gasification project sponsored by a consortium headed by American Natural Resources Company which aims to be the first commercial-scale synthetic gas facility in the United States.

Earlier this year the plant, which will be built in Beulah, North Dakota, appeared doomed both by the Reagan Administration's lack of enthusiasm, and by a court decision refusing the government permission to raise natural gas prices to cover the plant's construction costs. Riding a wave of congressional support, however, this plant has now been promised loan guarantees of \$2,000 million. The two other subsidies approved by Mr Reagan are \$1,100 million for an oil shale project jointly financed by the Tosco Corporation and the Exxon Corporation in Colorado, and \$400 million for another shale project in the same state sponsored by the Union Oil Company.

David Dickson

CORRESPONDENCE

Stanford's patent

SIR — Nowhere in your discussion of the Cohen-Boyer patent to Cohen and Boyer (*Nature* 13 August, p.571-574) do you give its number, which is 4,237,224.

In what you call the antecedents of the Cohen-Boyer patent the patentees included the work of Hershfield *et al.* at that time in the press but later in prosecution cited by the Examiner from *Proc. natn. Acad. Sci. U.S.A.* 71, 3455 (1974). The US Examiner also cited the earlier work of Chakrabarty to be found in his important US patent assigned to General Electric Company, No. 3,813,316.

It is essential in any discussion of possible infringement of 4,237,224 to look at the precise language of the 14 claims *per se* and not to rely on loose remarks on what commentators think they may or may not protect.

FRANK W. COUSINS

London SW1, UK

Unnecessary schism

SIR — I think that the Devil tries his best to drive the Church and scientists apart by splitting those who see truth in Darwin's theory of "Survival of the fittest" and the creation of the Universe by God. I believe the two go well together. If we think of "fitness" as being pleasing to God, who told the animals to "go forth and multiply", then this ties in well with "for the Lord watches over the way of the righteous, but the way of the wicked will perish" — Psalm Iv.6.

I believe that God is at work just as much in the Miller experiment as in the day of the primordial soup.

When the bible says that the Lord created the world in six days, we need not take this to mean 6×24 hours. In English we often use the word "day" to mean an indefinite period.

When Fox (*Nature* 6 August, p.490) suggests a theory of how nucleic acids are formed this reinforces my faith in God. Christians do not need to hide in the unexplained to see God at work but rather see him as all the more glorious when we see how he does it.

The gaps of knowledge suggest to me that the Creator is cleverer than all the world's scientists and all the world's computers put together.

SEAN JACKSON

Trinity College, Cambridge, UK

Discovery unmasked

SIR — Examination with modern microscopes of van Leeuwenhoek's original late-seventeenth century specimens and sections preserved amongst the manuscripts at the Royal Society is a most exciting project. However, Brian Ford (*Nature* 30 July, p.407) is mistaken in claiming that they have only just been discovered by himself. Clifford Dobell (1886-1949), who carried out his pioneering study of the Leeuwenhoek collection at the Royal Society throughout the 1920s, clearly states in his scholarly and indispensable *Antony van Leeuwenhoek and his "Little Animals"* (p.333) published in 1932 (and reprinted in paperback by Dover Books in

1960) that "Leeuwenhoek was one of the first — if not the very first — to study the structure of solid opaque bodies by means of sections. Some which he cut with his own hand by means of a sharp shaving razor are still in existence. They were enclosed in a little packet affixed to an early letter (Letter 4, 1 June 1674. To Oldenburg. MS. Roy. Soc. . .), and have remained intact to the present day".

Professor F.J. Cole, who justly stated "No student of Leeuwenhoek can fail to be deeply impressed by Dobell's classic monograph", also refers briefly to the existence of the specimens at the Royal Society in a publication in 1937 on "Leeuwenhoek's Zoological Researches". No doubt because it was published after Dobell's book, Cole's study, which appeared in two parts (*Ann. Sci.* 2, 1-46, 185-235; 1937), seems to have been generally neglected. It can be recommended as a most useful guide to Leeuwenhoek's letters, as a thorough study of Leeuwenhoek's histology, and particularly valuable because Cole has compiled a 37-page analytical index of tissues and specimens studied by Leeuwenhoek. The way in which Leeuwenhoek prepared his razor is described by him in two letters of September and November 1709 (*Phil. Trans.* 26, 493-502; 1709). Cole sums up Leeuwenhoek's histology thus: "He studied rough hand sections . . . never evolved a technique which would have enabled him to prepare sections of soft tissues without previously drying the material. Only near the end, in 1714 does he mention any method of staining . . . He first mentions section technique in 1674 and some of the sections he then cut with a 'sharp shaving razor' still survive attached to Letter 4 [1 June 1674]".

R. DEREK WOOD

Bromley, Kent, UK

but . . .

SIR — My contribution to your columns did not make the claim to which R.D. Wood objects. It actually began with the words: "The original specimens sent by the 'father of microscopy', Antony van Leeuwenhoek, are still in existence". The words "After not being seen since 1674 . . ." were added by *Nature* staff in processing the article.

Even so, this exciting discovery of Leeuwenhoek's fine specimens in pristine condition after three centuries or more is not dignified by citing occasional individuals half a century ago who noted the existence of the packets, but did not investigate what they contained. Contrary to what your correspondent implies, for the majority of the specimens there are no known historical records since Leeuwenhoek's time.

The value of these specimens as a source of information on Leeuwenhoek and his work lies in the remarkable fact that they have remained apparently undisturbed until I discovered them earlier this year¹. The historical record substantiates my belief: in particular, Dobell's remarks about the packet of 1 June 1674 are a helpful corroboration that the specimens remained (in his words) "intact to the present day".

BRIAN J. FORD

Cardiff, Wales, UK

1. Ford, B.J. *Not. Rec. R. Soc.* 36 (1), 37-59 (1981).

Matter of principle

SIR — Anthony Flew in his formulation of the principle of natural selection as non-random survival (*Nature* 16 July, p.192) claims to have denied this principle the tautological status some have accorded it. It does not seem, however, that this formulation circumvents the main problem — namely lack of empirical content of the principle. Non-random survival is an empirically empty concept if the range of possibilities the non-random survivors were supposedly selected from is unknown. The range of possibilities is known in a few instances only, where polymorphisms determined by multiple alleles correlate with environmental variables, for example, as in the industrial mechanism of the moth *Biston betularia*.

As a consequence natural selection can never have a broadness of empirical base commensurate with its claims for generality — a point which ought not to be overlooked if biologists wish to lay claim to being creatures of reason rather than faith and thus maintain the distinction in modes of thought between the biologist and the creationist.

BARRIE PEARSON

Winfrith, Dorset, UK

Thoughts on Popper

SIR — Halstead¹ has mentioned Popper's recantation² of some of his mistaken ideas on evolution. The unmasking of some of Popper's apparent errors occurred as follows. In May 1977 my paper³ on "The testability of the role of natural selection within theories of population genetics and evolution" was widely circulated and received for publication. In that paper I explained at length why Popper's ideas on the testability of natural selection seem totally wrong. Popper's own recantation appeared first somewhat later in his Darwin lecture (Darwin College, Cambridge, November 1977). His brief arguments seemed to be as wrong as ever and were based on isolated instances of natural selection, instead of referring to some population-genetic theories that involve natural selection. The Darwin lecture also contained an apparent flaw as regards Popper's views on epiphenomenalism⁴ (these views were based on evolutionary considerations). Ruse had earlier criticized Popper's apparent misconceptions on evolution⁵, but further scrutiny was called for. Hence, in a recent paper⁶, following the earlier critiques by Lewontin and Ruse of Popper's views on evolution, I analysed at great length the likely nature of the "theory of evolution". In that paper I noted that "it would be wrong to think that Popper has correctly assessed what matters in evolution", and I tried to exhibit in great detail the apparent fundamental errors of many of Popper's ideas on evolution, although my paper is mainly constructive.

G.D. WASSERMANN

University of Newcastle upon Tyne, Newcastle, UK

1. Halstead, L.B. *Nature* 292, 403-404 (1981).

2. Popper, K. *Dialectica* 32, 344 (1978).

3. Wassermann, G.D. *Br. J. Phil. Sci.* 29, 223-242 (1978).

4. Wassermann, G.D. *Mind* 88, 572-575 (1979).

5. Ruse, M. *Phil. Sci.* 44, 638-661 (1977).

6. Wassermann, G.D. *Phil. Sci.* (in the press).

NEWS AND VIEWS

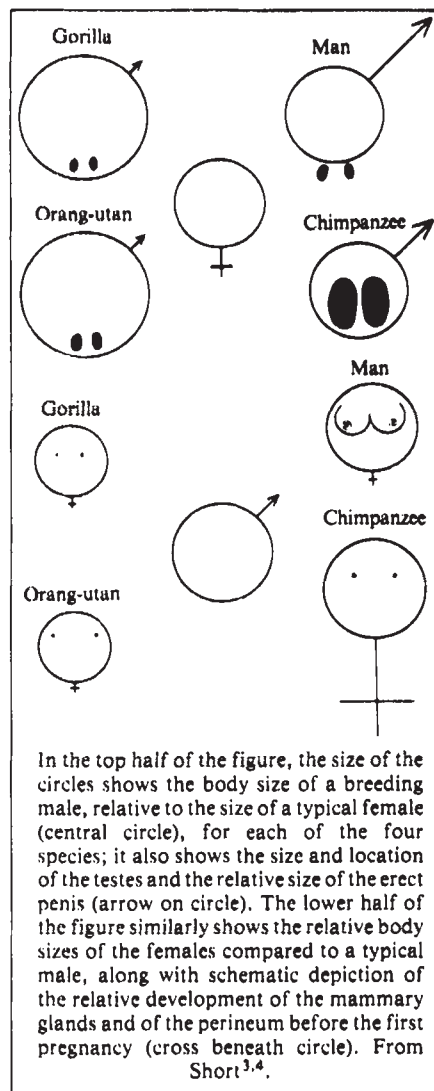
Outward signs of breeding

from Robert D. Martin and Robert M. May

ANIMAL behaviour, both within and between species, tends to exhibit much greater variability than does morphology or physiology. This makes it difficult to reconstruct phylogenies from behaviour patterns. The problems are particularly acute for mammals, where social behaviour can be very variable in its expression, depending on local ecological conditions and on prevailing demographic factors. Discussion of the origins of mammalian social systems therefore benefits greatly from any linkage to morphological or physiological features, which provide anchor points that are relatively stable in evolutionary terms.

A good example is provided by Harvey and colleagues' analysis of sexual dimorphism of body size among primates. Studies of living primates (and, incidentally, of other mammals) show that sexual dimorphism is absent in all species which exhibit habitual monogamy, whereas it may or may not be present in species with social systems where males have breeding access to more than one female (unimale or multimale polygyny). Accordingly, the apparent presence of sexual dimorphism in some of the earliest known (Oligocene) relatives of the Old World monkeys and apes from the Fayum deposits of Egypt suggests that they possessed polygynous mating systems of some kind².

Focusing on the Great Apes — orang-utan, gorilla, chimpanzee and man — Short has drawn together information about several morphological features that shed light on the relationships between sexual selection and behavioural ecology³⁻⁵. This example is especially interesting because the four species are genetically and biochemically very similar, yet phenotypically and behaviourally significantly different. The data assembled by Short are summarized schematically in the figure, which shows the sexual dimorphism in body size, the size and location of the testes, the relative size of the erect penis, and the relative development of the



mammary glands and the perineum, for each of the four species.

As discussed by Short, these morphological characteristics correlate well with the breeding systems observed for orang-utans, gorillas and chimpanzees (we will return to man below), and these in turn may be plausibly associated with the foraging behaviour of the females of the various species⁶.

Orang-utans are large arboreal frugivores, and the density of fruiting trees or

those with edible leaves and bark is not particularly high in the thick forests these creatures inhabit. Female orang-utans thus live in isolation and males occupy large 'core areas' which typically contain several females and from which they exclude other adult males. The mating system is thus basically a unimale polygynous one (effectively a harem group). There is consequently selection on the male for the physical size and strength to defend his core area, resulting in a pronounced sexual dimorphism where male orang-utans are about twice the size of females. On the other hand, copulation is a relatively rare event⁷; each of the few adult females within earshot of a given male is likely to come into heat only once or twice in 6 years or so (the mean birth interval is about 5-7 years). At such low copulation frequencies, there is no need for a high spermatogenic capacity — hence the relatively small testes. These circumstances, together with the lack of visibility in the forest canopy, put little selective pressure on sexual advertisement. Noting that the erect penis is about 4 cm long, Short⁵ adds the engaging observation that: "To what extent, if any, it is used in male-female display is uncertain, but its length certainly allows the animal to adopt a wide variety of copulatory positions, necessitated by the fact that copulation usually occurs when both animals are hanging from branches".

Gorillas are predominantly terrestrial, and feed on rather low-quality herbage which is usually abundant but in fairly widely dispersed patches. The females are thus relatively less concerned about competition for food and travel in small groups. Males compete for exclusive ownership of such groups (which fragment on the death of the dominant male). This polygynous, unimale breeding system is basically similar to that for the orang-utan, even though the ecological details are quite different. Again there is strong sexual dimorphism, little sexual advertisement (the erect penis is about 3 cm long) and small testes. Indeed, a typical male gorilla has three to four females in his troop, each having roughly 4 year intervals between births, so that he will encounter an oestrous female on average but once a year⁸; relative

Robert D. Martin is Reader in Physical Anthropology at University College, London, and Robert M. May is Professor of Biology at Princeton University.

to body weight, the gorilla testis is the smallest of any known primate.

Chimpanzees are terrestrial and arboreal omnivores, and live in relatively open habitats where vision is more important. However, the availability of their food is relatively low, so each female needs a large core area; thus, rather like the female orang-utan, the female chimpanzee tends to remain apart from the male. It is infeasible for a single male to defend a territory that embraces several females and the outcome is a system in which groups of males cooperate to defend a large area containing many females. In this system, there is little aggression among males within the group, which may explain the low degree of sexual dimorphism. Such social cooperation carries the consequence that males share the sexual favours of females that come into oestrus; chimpanzees have a promiscuous, multimale breeding system. Estimates⁵ of the size of the community of females available to a male, coupled with estimates of the frequency of oestrus for a given female, suggest that intercourse is almost a daily occurrence for a male chimpanzee, in vivid contrast to the gorilla and the orang-utan where it is roughly an annual event. The consequent emphasis on testis size, and sexual advertisement in both male and female, are as depicted in the figure.

One unsatisfactory aspect of the above story is that the comparisons are among so few species. In this issue of *Nature* (p.55), Harcourt and co-workers present a very nice study of the relationship between size of testes and type of mating system for some 33 species of monkeys and apes. As with recent studies of sexual dimorphism, an effective comparison of testis size among so varied a group of primates can only be conducted after the scaling influence of body size itself has been taken into account⁹. This reflects a theme in much of developmental biology: mechanical constraints on the way living machines can be constructed will tend to give general 'allometric' laws governing the way particular characteristics scale with body weight; superimposed on these broad patterns will be variations arising from adaptation to particular ecological or behavioural circumstances. Thus, for mammalian reproduction in general, it has recently been demonstrated that many different aspects of fertility (litter size, litter frequency and so on) exhibit systematic dependence on environmental circumstances (tropical versus temperate; tree-dwelling versus burrowing) once the systematic effects of body weight are removed¹⁰.

Harcourt *et al.*'s analysis confirms the suggestion by Short³⁻⁵ that, once body size has been properly taken into account, primate species with multimale breeding systems consistently have larger testes than those with unimale breeding systems (monogamous pairs or harem groups). As explained more fully by Harcourt *et al.*,

this is as expected: a greater capacity for sperm production will have obvious advantages wherever males may encounter direct competition within their social group for breeding access to females. It therefore appears that the relative size of the testes is a very useful guide to the basic breeding system of any primate species, reflecting fundamental adaptation regardless of minor adjustments of social behaviour to suit local ecological conditions. Extension of these comparisons to mammals generally should yield further useful insights, and a more reliable base for interpretation.

As in any approach dealing with only a single parameter relative to body size, there are complications, the most obvious of which is seasonality. For any given breeding system, males of a species which exhibits mating for only a restricted part of the year might be expected to possess larger testes than males of species breeding throughout the year. In addition, testis size is known to vary over the year in seasonally breeding mammal species. Any such seasonal effects must evidently be taken into account for a really clear evaluation of the implications of relative testis size. For example, among callitrichids it is known that the cotton-top tamarin (*Saguinus oedipus*) exhibits a marked seasonal breeding peak, whereas the common marmoset (*Callithrix jacchus*) breeds throughout the year with no obvious seasonal pattern¹¹. This might explain why cotton-top tamarins have relatively larger testes than common marmosets, despite the fact that both species are classified as monogamous. Similarly, Harcourt *et al.* point out that *Macaca* species, which generally show some kind of seasonal breeding peak, exhibit the largest relative testis size among all the primate species included in the comparison. There is also an apparent enigma in that the squirrel monkey (*Saimiri sciureus*) is known to exhibit very pronounced seasonality of breeding and is also classified as possessing a multimale breeding system, yet its testes (relative to body size) appear to be only moderately developed and are in fact both relatively and absolutely smaller than in the monogamous cotton-top tamarin. However, in this case it is possible that the data used by Harcourt and colleagues, which were derived from a laboratory colony, do not accurately reflect the natural situation. For this reason alone, it is unfortunate that no data on prosimian primates could be included in their analysis, since some of the more extreme forms of seasonal reproduction known among the primates are found in the Madagascar lemurs, with some species exhibiting mating for only a very few months each year¹². It can be predicted that among lemurs, testis size (relative to body size) should be larger than in monkeys and apes generally, but that monogamous species such as the indri (*Indri indri*) should show relatively small testes compared with

other lemur species, such as the ringtail (*Lemur catta*) and the sifaka (*Propithecus verreauxi*), which live in multimale breeding units.

Reliance on testis size as an indicator of breeding systems also begs the question of the timing of mating relative to ovulation. There is some evidence from baboons that dominant males mate with females closer to the time of ovulation than do subordinate males¹³, and it is not at all clear whether increased semen volume or better timing of mating would offer the better chances of breeding success for a particular male in a multimale breeding group. For example, if male A were to mate with a given female 48 hours before ovulation, while male B were to mate only 24 hours before ovulation, it is not known, at present, which male's spermatozoon would fertilize the egg.

As noted by Harcourt *et al.*, a male's success in competing with other males for mating with an oestrous female will not depend only on his sperm production capacity, as reflected by testis size. Increased sperm storage capacity will also enhance a male's chances of success in mating competition and one might therefore expect primate species with multimale breeding systems to exhibit a larger relative size of the epididymis than species with unimale breeding systems. Data are not yet available on epididymis weights in primates, but this is an obvious priority for future investigation.

What does all this tell us about ourselves? Man is unusual in that every mating system known for the order Primates as a whole can be found among contemporary human societies. Even if attention is restricted to those simpler existing societies that are uncontaminated by Western culture, the evidence remains equivocal: an analysis¹⁴ of 185 such societies found 74 per cent to be basically polygynous, although economic considerations and a shortage of women meant that in practice about half adopted monogamy. Indeed, some argue that human mating systems do not have a biological basis, but rather the form adopted in any given society is dictated largely by socioeconomic factors. On the other hand, the popular literature on human behaviour shows well established support for the notion that there is a biological basis for monogamy in *Homo sapiens*, and elaborate scenarios have been constructed around this idea.

The analysis conducted by Harcourt *et al.* shows that human testis size certainly does not accord with the range of values to be expected for a multimale breeding system. This line of evidence, however, does not discriminate between monogamy (whether serial or lifelong) and polygyny (unimale harem groups); relatively small testis size is found for monogamous primates such as the gibbon, and also, as illustrated in the figure, for polygynous primates such as the gorilla and the orang-utan.

Two other lines of evidence suggest, however, that our hunter-gatherer ancestors were likely to have been polygynous rather than monogamous. First, modern *Homo sapiens* exhibits sexual dimorphism both in body size (albeit to a mild degree, with men about 20 per cent heavier than women at any given age) and in form⁵. This contrasts with the well established rule that sexual dimorphism is not found among monogamous mammals. Second, as has been discussed with specific reference to the red deer¹⁵ (*Cervus elaphus*), it is to be expected that parents will allocate a greater proportion of their resources to sons rather than to daughters only in cases where "reproductive success varies more widely among males than females". For human gestation, it is known that more resources are devoted to a male fetus than to a female one, resulting in significantly higher birth-weights for male infants¹⁶. Since there is no cultural influence which might affect the developing fetus differentially according to its sex, this may be taken as indicating some biological basis for a mating system in which human males would exhibit more variation in reproductive success than would females; that is, polygyny or promiscuity rather than monogamy. Setting aside the trends in degree of sexual dimorphism and relative weight of testes, the other patterns in the figure are mainly concerned with the conspicuousness of the genitalia (size of erect penis, whether or not the testes are pendulous and so on). These patterns have not yet been compared systematically among primate species, and the evidence they offer about mating systems can be subjected to an indecisive variety of interpretations.

Note the special importance of the study by Harcourt *et al.* in these attempts to reconstruct the breeding systems of our human ancestors by broad comparisons among primates. Earlier work tends to rule out monogamy, but makes little distinction between multimale and unimale breeding groups; the current study of the relative size of testes now tends to exclude multimale

systems. Monogamy may, of course, nonetheless be the optimal system in certain modern societies. Much work on birds, mammals and other creatures indicates that monogamy often emerges when large investments in the offspring necessitate cooperation between the parents. It is possible that the increasing dependence of the human infant, associated with progressively increasing brain size and cultural complexity, has favoured culturally deter-

mined monogamous tendencies in various human societies, even though these have not led to the complete suppression of biological indicators of a polygynous ancestry.

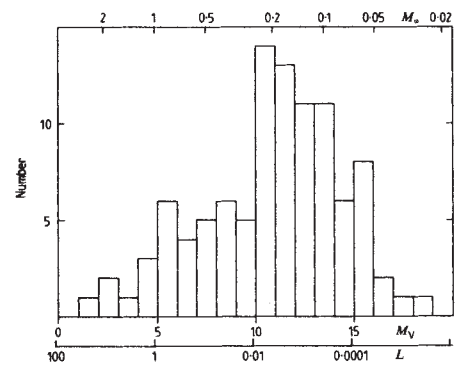
In brief, it seems likely that our morphology is that of a species with unimale breeding groups but that these biological antecedents are today often overlain by extremely powerful socio-economic determinants. □

The least luminous star

from David W. Hughes

ASTRONOMICAL records are always interesting and it is a matter of considerable excitement when one is broken. I. Neill Reid of the Department of Astronomy, University of Edinburgh, and Gerard Gilmore of the Royal Observatory, Edinburgh, recently reported in the *Monthly Notices of the Royal Astronomical Society* (196; 15P, 1981) that a present record holder, star VB10 in Aquila, has probably been overtaken by star RG0050-2722 in Sculptor. The race is to see which one is the least luminous. VB10 has an absolute visual magnitude, M_v , of +18.57; RG0050-2722 has an M_v of about +19. (M_v is equated to the apparent magnitude a star would have if it was 10 pc from the observer. For example, the Sun has an apparent magnitude of -27 and an M_v of +4.83.)

The authors have been studying the stellar distribution in the region of the South Galactic Pole using the UK Schmidt Telescope. Photographic plates were taken with a range of filters, typical examples being the B, V, R, I, J, H and K filters in the visual and near infrared. Colour relationships between the apparent magnitudes of RG0050-2722 in each specific wavelength band led to an estimation of the star's spectral type and luminosity class. These data fix a position for RG0050-2722 in the Hertzsprung-Russell diagram from which it is possible to read off the star's absolute magnitude. Using three different filters, M_v was estimated to be 20.5 ± 1 , 19.0 ± 0.5 and 19.5 ± 1 . From the apparent magnitude of the star, its distance was estimated to be 25 ± 6 pc. RG0050-2722 is the first star of very low luminosity to be discovered by purely photometric criteria. It is classified as being a metal-rich dwarf and as such, Reid and Gilmore assume that it is on the main sequence. It seems that TiO in its atmosphere provides extensive blanketing in the 0.7 - $1 \mu\text{m}$ spectral region. The



Histogram showing how the 100 stars closest to the Sun are distributed according to their absolute visual magnitude, M_v , their luminosity, L (in units of the solar luminosity, 4×10^{33} erg s⁻¹) and their mass, $M_\odot$ (in units of the solar mass, 2×10^{33} g). The data have been taken from C.W. Allen *Astrophysical Quantities*, 3rd edition.

spectral energy distribution indicates that the star has a surface temperature of $2,625 \pm 100$ K.

Stars of such low luminosity seem to be rare. The figure shows the distribution of stellar numbers as a function of absolute visual magnitude. It includes the 100 closest stars and, considering the capabilities of the largest telescopes, it can be concluded that very little has been overlooked. There is a reasonably sharp upper and lower mass cutoff. The upper limit of the luminosity function is about $M_v = -8$ and the equivalent maximum stellar mass is about 100 solar masses ($M_\odot$). As can be seen from the figure, no star that massive is near the Sun. Reddish (*Sci. Prog.* 50; 235, 1962) estimated that the lower stellar mass limit is $0.027 M_\odot$.

If we assume that RG0050-2722 obeys the normal mass-luminosity relationship, its mass can be calculated to be $0.023 M_\odot$ which puts it very close to the lower limit. Reddish found that the average stellar mass was $0.11 M_\odot$.

Where do stars end and planets begin? Jupiter, the largest planet in our Solar

1. Harvey, P.H., Kavanagh, M. & Clutton-Brock, T.H. *J. Zool. Lond.* 186, 475 (1978).
2. Martin, R.D. *Nature, News & Views* 287, 273 (1980).
3. Short, R.V. *Adv. Study Behav.* 9, 131 (1979).
4. Short, R.V. *J. Reprod. Fert. Suppl.* 28, 3 (1980).
5. Short, R.V. in *Reproductive Biology of the Great Apes* (ed. Graham, C.E.) 319 (Academic, New York, 1981).
6. Wrangham, R.W. *Soc. Sci. Inf.* 18, 335 (1976).
7. Galdikas, B.M.F. in *The Great Apes* (eds Hamburg D.A. & McCown, E.R.) 195 (Benjamin, 1979).
8. Harcourt, A.H., Fossey, D., Stewart, K.J. & Watts, D.P. *J. Reprod. Fert. Suppl.* 28, 59 (1980).
9. Martin, R.D. *Z. Morph. Anthropol.* 71, 115 (1980).
10. May, R.M. & Rubenstein, D.I. in *Reproductive Biology*, Vol. 6, 2nd edn (eds Austin, M. & Short, R.V.) (Cambridge University Press, in the press).
11. Brand, H.M. *Lab. Anim.* 14, 301 (1980).
12. Van Horn, R.N. *Prog. reprod. Biol.* 5, 181-221 (1980).
13. Rowell, T.E. in *Social Communication among Primates* (ed. Altmann, S.A.) 15 (University of Chicago Press, 1967).
14. Ford, C.S. & Beach, F.A. *Patterns of Sexual Behaviour* (Eyre & Spottiswood, London, 1952).
15. Clutton-Brock, T.H., Albon, S.D. & Guinness, F.E. *Nature* 289, 487 (1981).
16. Gibson, J.R. & McKeown, T. *Br. J. soc. med.* 6, 152 (1952).

David W. Hughes is in the Department of Physics, University of Sheffield.

System, has a mass of $0.0010 M_{\odot}$ — about one-twentieth the mass of RG0050-2722. Is the mass function such that there is a collection of objects, as yet undiscovered, with masses between the two? Probably not. When considering stellar formation, the minimum star size is governed by the size of the smallest proto-star cloud fragment which is stable against disruptive forces. This probably depends on the rate of loss of angular momentum, the opacity of the fragment and the properties and quantities of dust in the interstellar cloud. Planetary formation, on the other hand, depends on the size, turbulence, density, composition and temperature of the flattened nebula disk around a new star. The size distribution of the resulting planets probably differs completely from the size distribution of the low-mass stars.

However, their compositions are probably very similar.

Kumar (*Astr. J.* **69**; 143, 1964) suggested that stars with masses of less than $0.08 M_{\odot}$ could not produce energy by the nuclear transformation of hydrogen into helium, as do main-sequence stars, because their central temperature was never high enough. He proposed that, as the low-mass stars contracted, they would fall straight past the main sequence and go directly to a cold 'black dwarf' stage. If this is the case, it raises problems for the photometric parallax method used by Reid and Gilmore in the low-luminosity region of the Hertzsprung–Russell diagram. Kumar infers that there may be 200 such 'planetary' bodies within 5 pc of the Sun, in a volume containing about 50 stars. None has yet been found. □

The cloning revolution meets human genetics

from Bob Williamson

THE Sixth International Workshop on 'Human Gene Mapping' held in Oslo in June was a small conference attended by both established experts and younger scientists new to the field. During the eight years since the first workshop, the number of chromosomal assignments of human genes has grown from approximately twenty to several hundred, with about one hundred of them precisely mapped to a chromosomal location.

What does 'precisely mapped' mean? This point was discussed by McKusick (Johns Hopkins University), Southern (University of Edinburgh) and Skolnick (University of Utah). There are two sorts of map — a phenotypic map, where characteristics (such as an enzyme activity or an inherited disease) are mapped to single stained human chromosomes, and a molecular map, where the organization and sequence of the DNA composing the genes is determined by the molecular biologist.

The distances used by the two groups of mapmakers differ. Banding techniques, in which the chromosome structure is stained with acidic or basic dyes (the functional significance of the banding is still not clear in mammals), enable a good cytogeneticist (on a good day) to distinguish 20 or so bands, perhaps a few more on extended chromosomes. As an average-size chromosome (such as the X chromosome) contains approximately 100,000,000 base pairs (bp), the nearest gene assignments

using banded chromosomes are still about 5,000,000 bp apart (5,000 kb, Kilobase pairs).

In contrast, the molecular biologist studies individual gene sequences cloned as recombinants and can determine a sequence of DNA several hundred or thousand base pairs long. The biggest gene fragment that can be cloned (in a cosmid) is only 50 kb, one-hundredth of the cytogeneticist's best resolution. Excitement mounted at the meeting as it became apparent that this enormous gap between the distance looked at by the traditional human geneticist and by the molecular biologist will be bridged during the next two years.

The classical methods of studying genetic linkage — comparing the inheritance of a disease with that of a protein marker within a family, or looking for the expression of a protein by a somatic cell hybrid between mouse and human cells containing only a limited number of human chromosomes — only work when a family, or a cell line, exists with just the right combination of genetic traits. Even then, the gene assignment often involves a cytological analysis, and therefore is rarely more accurate than $\pm 5,000$ kb, unless the genes are extremely closely linked. Polymorphisms are required in the family analysis, but not when genes are expressed in hybrid cells.

The new molecular techniques for determining linkage use cloned DNA sequences, and look for polymorphisms where small or single-base changes occur in the DNA sequence between two homologous chromosomes in an individual. As first pointed out by Jeffreys (University of

Leicester, UK), the polymorphisms occur about once every 100 or so base pairs, particularly in the more than 90 per cent of the DNA that does not code for protein, and appear to be random and non-selected and therefore can be regarded as completely neutral mutations. They can be followed through several generations and are inherited in mendelian fashion.

There are only about 50 cross-overs, or chromosomal exchanges of genetic material, during a single meiosis, and very rarely will a single gene be involved. Because of this, Skolnick and others calculated that it should be possible to map the entire human genome using a relatively small number of polymorphic DNA fragments cloned in phage or plasmid vectors. The only requirements for such an exercise are that meiotic cross-overs are relatively rare events, the clones are randomly distributed, the polymorphisms can be followed through a family and they can be linked to some phenotype, whether a hereditary disease or the expression of a particular protein or nucleic acid.

Several laboratories (University of Utah, Johns Hopkins University, Harvard/MIT, St Mary's London and University of Edinburgh) are well on the way to constructing random human clone libraries which can be used for gene mapping, at least for some chromosomes. The techniques for chromosome isolation vary from the use of rodent–human hybrid cells (University of Utah) to the use of the fluorescence-activated cell sorter to sort chromosomes (St Mary's, London, Beatson Institute, Glasgow). Once a chromosome-specific probe has been isolated, it can be mapped using translocations or chromosome fragments, or by *in situ* hybridization, and used in turn to map any phenotype by linkage inheritance studies. An example of the use of *in situ* hybridization for gene localization was described by Ferguson-Smith (University of Glasgow), who collaborated with Malcolm (Queen Elizabeth College, London) to use a tritium-labelled probe for a V-region gene of immunoglobulin light chain and showed it resides close to the centromere on the short arm of chromosome 2.

In future linkage studies, a family with a disease will donate small blood samples, from which leukocyte DNA samples can be obtained, and the laboratory determine whether there is a co-inheritance of a DNA sequence (as measured by the size of a fragment seen on a Southern blot autoradiograph with a specific cloned gene probe) with the phenotype (in this case, the disease). It is probable that by the next workshop in two years' time, these techniques will allow the mapping of any phenotype which is defined by a single gene to a single chromosome locus, particularly if the correct chromosome is already known.

Southern pointed out that it is then merely a question of walking the genome

Bob Williamson is Professor of Biochemistry at St Mary's Hospital Medical School, University of London.

How 'unique' can meteorites be?

from Robert Hutchison

ALL meteorites may be considered 'unique' with respect to their time of fall, which determines when the radionuclides produced by cosmic radiation begin to decay and when terrestrial weathering sets in. However, 'uniqueness' of meteorites is usually defined in terms of chemical composition and mineralogy, although oxygen isotopic ratios may be significant. No meteorite has properties which suggest that it came from outside the Solar System.

The most recently fallen meteorite to qualify as 'unique' landed near Acapulco, Mexico, on 11 August 1976. The single, 1.9 kg stone was extensively studied by Palme and colleagues in Mainz and Paris (*Geochim. cosmochim. Acta* 45; 727, 1981). Like members of the H-group of chondritic meteorites, it has 22.7 wt per cent metal and 3.6 per cent sulphide (the remainder being essentially silicate), but differs from them in that it lacks chondrules and is coarse grained. A more fundamental difference lies in the oxidation state of Acapulco, which is intermediate between that of the highly reduced enstatite chondrites and the more oxidized H-group.

In major element chemistry Acapulco closely resembles the H-group, but the careful analyses undertaken at Mainz show that it differs in its content of some minor and trace elements. It has three times more phosphorus than have H-group chondrites; the mineralogical work, carried out in Paris, attributes this to a high abundance of chlorapatite, a calcium mineral known to accept uranium, and which is an order of

magnitude more abundant in Acapulco than in H-group chondrites. These and other chemical properties of the meteorite suggest that the mineral assemblage has undergone some fractionation by partial melting. Enrichment in the refractory siderophile elements rhenium, osmium and iridium is taken as an indication that pre-accretionary, nebular processes also played a part in determining the chemical composition of Acapulco. But herein lies a problem. The chemistry and mineralogy of the meteorite seem to have been determined first by nebular condensation, then by accretion and igneous processes. During the igneous event it is estimated that the temperature was about 1,100°C, yet the planetary noble gases ^{36}Ar , ^{84}Kr and ^{132}Xe were not driven off. The K-Ar age of 4.7 ± 0.3 Gyr indicates an uneventful history thereafter, until the potential meteorite was broken from its parent body during a collision some 5 Myr ago, when it became exposed to cosmic radiation. The cosmic-ray exposure age is based on the contents of the spallogenic isotopes ^3He , ^{21}Ne and ^{38}Ar . Histories such as this are not uncommon among stony meteorites.

Chemically and mineralogically the Acapulco meteorite is certainly unusual and interesting, but is it really unique? Palme and his co-workers identify six meteorites, plus silicate inclusions in an

iron meteorite, which have chemical affinities both to Acapulco and to H-group chondrites. The only meteorite of the six which has chondritic structure, Kakangari, is itself unique in several respects (*Nature* 251, 128; 1974; 265, 230; 1977; and *Earth planet. Sci. Lett.* 40, 168; 1978). The remaining meteorites or silicate inclusions differ from each other in their detailed chemistry (such as total Fe contents and/or Al/Si ratios), but igneous differentiation can conveniently account for differences other than in oxidation state. The oxygen isotopic data of R.N. Clayton and co-workers seem to provide the best test of inter-relationships. Acapulco is apparently related to the 'unique' stony-iron, Lodran, although the latter is richer in metal and greatly impoverished in aluminium. Finally, Acapulco most closely resembles an Antarctic meteorite, Allan Hills 77081. The Antarctic stone has over twice as much sulphide as Acapulco, but in other chemical and mineralogical properties the match is excellent. Oxygen isotopic data were not available for Allan Hills 77081.

Technically, we may conclude that Acapulco is different from all other known meteorites, although the difference between it and Lodran or Allan Hills 77081 may be compared with that between neighbouring rocks from the top of the Earth's upper mantle. Philosophically, it is by studying rare or 'unique' meteorites that we increase our knowledge of the spectrum of planetary compositions and of the processes which took place on minor planetary bodies.

Robert Hutchison is at the British Museum (Natural History), where he is in charge of the National Collection of Meteorites.

towards the determinant that actually causes the phenotype — a linkage, of course, may be a long way away, but new techniques such as those described originally by Goss and Harris (University of Oxford), where chromosomes are caused to fragment after irradiation, or 'microcells' and cells transiently incorporating cloned DNA segments, which were discussed by Ruddle (Yale University), can be used to bridge the gap between the 50,000 bp clone and the 5 megabase pair chromosome band. Unlike traditional 'gene walking' methods, these new techniques ('jogging the genome') may allow a chromosome to be roughly sequenced in a few months, using random clones and a computer.

When a total gene map exists, McKusick predicted that it will be possible to carry out 'reverse diagnosis', from gene to phenotype. This will be particularly useful in the diagnosis of dominant diseases, such as Huntington's chorea, where persons at risk live with anxiety for many years before finding out whether the disease develops in middle age.

A clear example of the remarkable advantages of combining cellular and molecular analysis was presented by Francke (Yale University). Using a human insulin gene clone provided by Rutter's laboratory (University of California, San Francisco), she showed that this gene is on the short arm of chromosome 11, almost adjacent to the human β -globin gene. It may even be possible to diagnose sickle cell anaemia using a gene probe for insulin in the future! The insulin gene is of interest in another respect: it is one of the few genes (another on chromosome 14 was described by White, University of Utah) shown to have a polymorphic short sequence of DNA which can be inserted either in a variable number of copies, or in a number of discrete sizes, at a single position on the chromosome. This is another indication that there may be 'transposons' in the human genome, similar to those that have been found in bacteria and *Drosophila*.

Evolution can also be studied by gene mapping. Not only do primates and man share chromosome assignments, but these often hold true when comparing man with

mouse, kangaroo and, in some remarkable cases, even the fruit fly. Attardi (California Institute of Technology) gave a comprehensive account of the structure of the mitochondrial genome (the twenty-fifth chromosome, or 'little brother'), showing not only its extreme economy of function, but also the way in which it gives clues as to the evolution of living things, particularly by the study of the relatively unsophisticated mitochondrial tRNAs.

The meeting was supported by the World Health Organization as well as by the United States 'March of Dimes', and it is very encouraging that the WHO is beginning to take a major interest in the control of human genetic disease. Genetic diseases are a major cause of ill health in developing countries, and Kuliev (University of Moscow), who has recently taken responsibility for the WHO Genetics Programme, has been approached by many health workers from developing countries encouraging WHO to start a world programme to control genetic diseases, especially those affecting maternal and child health. □

Corals — the reef and man

from P.K. Swart

CORAL reefs have long been recognized as ecosystems of special beauty and diversity. Also, many millions of people depend for their existence on reef systems, particularly in the Indo-Pacific region. The interaction between conservationists and those concerned with the economic importance of coral reefs provided the background for the Fourth International Coral Reef symposium entitled 'The Reef and Man', held in Manila, the Philippines, from 18 to 22 May 1981.

The location of the conference, in the Philippines, marks an awareness of how important reefs are in the economy of this country. The problem to be solved is how man can live in harmony with the reef — both maintaining the stability and diversity of the reef environment and providing a livelihood for the surrounding populations.

A major problem facing reefs in many countries is their over-exploitation to provide curios and souvenirs for the Western tourist trade. The Philippines appears to be the largest supplier of coral and shells, exporting mainly to Europe, Japan and America (S.M. Wells, Species Conservation Monitoring Unit). The removal of coral not only destroys the reef but selective cropping of molluscs, such as the Triton, may have a role in the population explosion of coral predators.

There has been much publicity given over the past decade to the problem of *Acanthaster planci* (the Crown of Thorns starfish) and there is indeed no doubt that certain areas of the Great Barrier Reef are being subjected to renewed infestation. There is controversy over whether the animal is a specialized R-strategist whose population is out of control as a result of some as yet unidentified influence, conceivably brought about by man's activities, or whether the outbreaks are natural cyclical phenomena. A.M. Cameron (University of Queensland) argues that while cyclical outbreaks are common in herbivores, they are undocumented in carnivores. In light of the apparent problems of controlling *A. planci*, outlined by R.A. Kenchington (Great Barrier Reef Marine Park Authority), further research is needed to locate vulnerable areas in its life cycle and to establish whether outbreaks are indeed a cyclical phenomenon.

Perhaps the most confusing aspect of coral reefs is the apparent diversity of coral species. The phenotypic plasticity of corals has caused countless taxonomists to add to the confusion by creating many

synonymous species. At last some Australian taxonomists are recognizing the problems which face the non-taxonomist working with corals and eventually a usable key for all genera and species may be developed. A prototype for one of the more complicated genera was presented by C.C. Wallace (James Cook University, Australia).

Two aspects well covered in the conference were reproduction and coral interactions. Both may be important in shaping coral communities, although distinct doubt was cast on the role of coral-coral interactions by R.H. Bradbury (Australian Institute of Marine Sciences). He does not dispute that coral interactions occur, but the analysis of data collected by himself and others showed no evidence of the role of such interactions in shaping coral communities. However, his opinion is not held by everyone, even those working from the same institution.

Data presented on coral reproduction indicated variability in reproductive strategies by both sexual and asexual means. Asexual reproduction by fragmentation appears to be important in some species of *Acropora* and *Porites*, and a hitherto unknown method of asexual reproduction, termed 'polyp bale out', has been documented in *Stylophora pistillata* (P.W. Sammarco, Australian Institute of Marine Sciences). In this latter method, individual polyps appear to abort the original colony, each one forming a new clone.

The most significant phenomenon occurring in coral reefs, calcification, was the most ignored at this conference.

Despite the passage of over two decades since Goreau proposed a hypothesis for calcification in corals, little progress has been made on the crucial questions of how calcium is transported to the sites of calcification and, even more fundamentally, where these sites are. Similarly the role of the endosymbiotic dinoflagellates is still uncertain. A combined approach is needed to answer these questions as corals are both geological and biological entities, and all too often scientists in one discipline will disregard through ignorance evidence provided by the other science. For example, how are trace elements incorporated into the skeleton and are their concentrations different from inorganic precipitates? Some workers (B.E. Brown, University of Newcastle, and R.E. Dodge, Nova University) have reported that corals from apparently polluted areas have skeletal compositions similar to open oceanic ones, thereby casting doubt on the idea that corals accurately reflect oceanic chemistry.

Since the first Coral Reef Symposium held in Manapam, India, in 1969, the number and interests of reef workers have grown enormously. Consider alone the 290 papers given in four parallel sessions at this symposium as opposed to the 37 at Manapam. This growth of coral reef research has been reflected in the formation of the International Society for Reef Studies whose journal 'Coral Reefs' will appear next year. The coral reef community look forward to the Fifth International Symposium to be held in Tahiti in 1985. □



THE BRITISH ASSOCIATION

The Fifty-first Annual Meeting of the British Association was opened yesterday under the presidency of Sir John Lubbock, Bart., M.P., F.R.S., at York, the birthplace of the Association fifty years ago (September 27, 1831).

Many cities have received the Association twice, but few three times. York will now be one of the latter. The local committee have issued an extremely useful programme of their arrangements, which contains articles on the zoology, botany, and geology of the neighbourhood, and a description of the various excursions. An interesting article on "The York Founders of the Association" is contributed by Archdeacon Hey. An exhibition of art and industrial produce, and a collection of scientific apparatus, will be open during the week. Four excursions are

organised for Saturday, September 3: to Scarborough; to Castle Howard; to Helmsley and Rievaulx; and to Brimham Rocks and Harrogate. On the following Thursday there will be seven excursions: to Bolton Abbey and the Strid; to Cleveland; a coast excursion; to Gristhorpe, Speeton, and Scarborough; to Whitby; to Wensleydale; and to Aldborough and Boroughbridge. Among the more important manufactories which will be visited are the telescope works of Messrs. Cooke and Sons, the workshops of the North Eastern Railway, the York glass works, and some extensive confectionery works. Naturalists will be glad to learn that the county possesses a fauna which comprises 513 out of the 717 British Vertebrata, viz. 46 mammals, 307 birds, 12 reptiles, and 148 fishes. It also furnishes 71 per cent of the British flowering-plants and ferns. Geologically the county consists of rounded Chalk Hills, Oolite overlying the Lias, Trias covered with glacial drift and alluvial deposit, and a narrow band of Permian strata. Many opportunities will be afforded to members of studying the geology of the district.

From *Nature* 24, 401, 1 September, 1881.

The BA — and science — in retrospect

The afternoon of 2 September at this year's 150th anniversary meeting of the British Association at York was given over to a retrospective view of science. The following are extracts from the addresses at that symposium.

Medicines then and now

D R Laurence

WHILE it is certainly true that prevention is better than cure and it may also be true that the medical profession and society have given too little attention to, and invested too few resources in prevention, I have no doubt that enormous numbers of people will indefinitely continue to fall sick. We may well have neglected preventive medicine, but the sick person is unlikely to be immediately interested in that.

Medicines have always been a major vehicle for treating the sick and are likely to remain so. And a modern scientific/technological society will seek to make medicines more efficient by improving those it has and by finding new medicines.

Although I am a clinical pharmacologist, in principle, I am against medicines; we ought not to need them. In *theory*, nobody in his right mind takes into his body foreign chemicals (that are not nutrients) unless he has no option. But in real life we know it is very different. The great Sir William Osler, Professor of Medicine successively at universities in Canada (McGill), the United States (Philadelphia, Johns Hopkins) and at Oxford, wrote in 1894: "But know also, man has an inborn craving for medicine . . . It is really one of the most serious difficulties with which we have to contend".

Can medicines do good? Yes, you have all heard of penicillin. Can medicines do harm? Yes, you have heard of thalidomide. Do medicines do more good than harm? Yes they do, and the insatiable appetite of the public for medicines demonstrates that this is a view generally held, despite a vocal dissenting minority that tells us that the public is being conned by the medical profession and pharmaceutical companies.

With medicines, we are in the business of getting benefits of a peculiarly important, intimate, personal kind, without doing harm. The whole idea of risking harm when all we want is to escape discomfort, pain, misery or physical disability, is repugnant. We resent it, but if we demand the benefits and refuse to accept the risks, we refuse to accept reality. Such a refusal is alarming in supposedly mature adults, though I suppose it is human nature.

The major contribution of medicines is principally to the quality of life, though quite often they cure and in doing so save lives. Medicines are the very ideal of utility, which is specially gratifying to a member of University College, London — an institution founded shortly before the British Association by the utilitarian

philosopher Jeremy Bentham. He propounded the principle of utilitarianism — that a thing is to be judged for its utility according to its capacity to promote the greatest happiness of the greatest number. This is what drug science, pharmacology, is all about — the greatest 'happiness' or well-being of the greatest number.

But pharmacology is also the study of the effect of chemicals on living tissues, so if we are to discover and use medicines for maximum benefit with minimum risk, we must understand not only the substance but also the patient and the disease. So what was the position in 1831? Not good. Generally, a medicinal desert with a few oases — mercury for syphilis, quinine for malaria, digitalis for heart disease, opium for pain — discovered by chance and blind empiricism. Here is a characteristic passage on tuberculosis from a manual of therapeutics of 1832:

If . . . the symptoms announcing the first stage of phthisis show themselves . . . we must seek to lessen the pulmonary congestion by the moderate abstraction of blood . . . through dry cupping and bathing the feet in hot water, and by the use of demulcents, refrigerants and aperients; the cough may be quieted, in some degree, by opium . . . and sulphureous baths. Other remedies include ipecacuanha in emetic doses, inhalation of vapour of tar or ether.

After more of this the author concludes, "and when the larynx has become affected . . . we can only deplore the insufficiency of our art".

On 30 September 1846 in the Ether Dome at the Massachusetts General Hospital in Boston. "In the gallery, sceptical spectators gathered for the news had spread that a second-year medical student had developed a method for abolishing surgical pain". It was a success. The strong men to hold down the struggling patient were not needed. Within 4 weeks ether was used in London where the surgeon declared, "This Yankee dodge, gentlemen, beats mesmerism hollow".

Ether had first been made by empirical chemical experiment in 1540 and it is chastening to think of the intervening 300 years of agony under surgery as well as of the restriction of the range of surgery due to the imperative need for speed. With ether the surgeon had time for careful work, and with the advent of techniques excluding bacteria (asepsis) and antibacterial drugs to be used when infection was not avoidable, the wonders of modern surgery became possible.

We come, therefore, to the years around the turn of the century. In 1899 aspirin (acetylsalicylic acid) was introduced into medicine. This is a most extraordinary drug. It derives from observations that can only be described as a mixture of superstition and serendipity. In mid-eighteenth century England, after the introduction of Peruvian bark (which contained quinine) which was good for malaria, the bark of native trees was explored, especially those growing in damp and 'malarial' places. A clergyman tasted the bark of the willow tree and found it bitter, like Peruvian bark. The bitterness was due to the glycoside salicin which yields, following a series of reactions, sodium salicylate and eventually aspirin.

Aspirin soon became the standard home remedy for aches, pains and fevers. It is still important in rheumatoid arthritis and rheumatic fever. Following the brilliant analysis of its mechanism of action on a cellular enzyme (prostaglandin synthetase) that mediates tissue inflammatory and painful responses to injury, its use has now extended to the prevention of heart attacks by its action modifying the stickiness of the cellular element of the blood (platelets), the function of which is to stop bleeding from damaged blood vessels. Low doses of aspirin have been shown to reduce the incidence of second attacks of coronary thrombosis. Thus aspirin, after 80 years is still, as a result of scientific investigation of the biochemistry of the body, finding new uses. But the same mechanism of action is not useful in the stomach. In patients taking aspirin daily, there can be a steady daily loss of as much as 5 ml (one teaspoonful) of blood. Thus aspirin has benefits that can be valuable, and risks that can be serious, but as far as I know, nobody seriously wants to stop its sale to the public.

Soon after aspirin was introduced, perhaps the greatest medical scientist of all time, Paul Ehrlich, was making a scientific revolution. He wrote:

We must search for substances which have an affinity for the cells of parasites and a power of killing them greater than the damage such substances cause to the organism itself, so that the destruction of the parasite will be possible without seriously hurting the organism . . . we must learn to aim, learn to aim with chemical substances.

And Ehrlich did so, inventing the first cure for syphilis, salvarsan. This is what medicinal chemistry and pharmacology are about — learning to aim with chemical substances, the principle of selective toxicity. This principle is the basis of chemical environmental control. By extension we learn to aim at the function of different parts of the human body to get

solely the effect we want. We study, down to molecular level, the function of normal and diseased tissues, by finding out how they work biochemically, and then we make substances that selectively mimic or antagonize the chemical mediators of natural or diseased functions of the body.

Three other dates demand mention in this survey of 150 years:

- In the 1890s, following the early studies of the physiology of endocrine glands, therapy for thyroid gland deficiency (hypothyroidism) was started by oral administration of sheep's thyroid "lightly fried with currant jelly". This restored the wretched victims to complete health and normal life. Nowadays the pure hormone (thyroxine) comes as a small white tablet.

- 11 January 1922, when insulin (obtained from animal pancreas) was first administered to man. A diabetic physician who escaped an early death as a result of the discovery of insulin, wrote in the 1925 preface to his standard book *The Diabetic Life*: "Five years ago the *Diabetic Death* would have been a more suitable title for such a book as this . . . There is no reason now why a diabetic should not lead a long and normal life".

- In 1960, the thalidomide disaster dramatically revealed the full horror of what can happen when foreign chemicals are introduced into the body, and has been chiefly responsible for the massive systems of government regulation of new and old drugs now in operation. It is notable that in Britain the principal advisory body is named The Committee on Safety of Medicines, reminding us constantly of the chief stimulus to government intervention.

But why must we accept risks? Why cannot they be eliminated? One reason is that the chemical transmitters that mediate body function are released by nerve endings right at the cell designed to respond; their release is extremely precise and localized. This allows similar communication mechanisms to be employed in widely different systems. Important enzymes, though localized in cells, serve different functions in different tissues. When we swallow a medicine designed to affect say, inflammation in the joints, we flood the body with it. We lose selectivity. Although ingenious scientists are attacking this and other problems, we cannot soon expect drugs that will do only what is wanted and nothing else.

To anyone who demands totally safe drugs, or that drugs be proved to be totally safe before they are introduced into medical practice, I would say three things:

- You cannot have such safety, the option is not available, any more than it is available for cars, aircraft or domestic gas and electricity installations.

- Consider the consequences of turning your back on anaesthetics, antibiotics and, yes, on tranquillizers, as well as on modern transport, or domestic amenities.

- The whole of life is a benefit: risk state in which we seek to maximize the one and

minimize the other.

What then should a society, accepting the reality that medicines carry risks of even serious harm, say to the individual who has suffered that harm? In principle, society should care for its disabled members according to their need, and not selectively according to the cause of their disability. But society has shown its intention to be selective and to give extra support to cases it deems specially deserving — industrial injuries, criminal injuries and now injury due to vaccines when these are used according to government policy (but not otherwise). The notion that compensation is given for an injury due to an attempt to avoid disease and not where injury is due to the disease itself points to the essential irrationality of selective compensation.

If we are to compensate for drug injury, should we not compensate for injuries due to other forms of treatment (surgery) or even for the "accident" of disease? Piecemeal warm-hearted but unthinking demands for special treatment of special cases can lead to a hierarchy of disabilities in relation to the support they get. There will be anomalies and injustices, and bitterness on the part of those at the bottom of the heap — those born disabled with no attributable cause, for example. Is this what society wants?

Liability

Nevertheless there seems no doubt that society intends to have a special scheme for drugs. So what might be done? For drugs prior to licensing for general use, there should be strict liability on the manufacturer. For drugs licensed for general use there should be a no-fault scheme centrally funded by the parties involved.

The implementation of a sensible and manifestly fair scheme will do more to promote public understanding than any amount of exhortation. But a word of caution. There is a natural human tendency to seek to place blame for personal misfortune outside ourselves. All disability that occurs when taking a drug is not necessarily due to the drug. Ordinary fairness dictates that special compensation be given only in cases where the cause can be firmly established, because if this is not done we shall have created what is hardly more than a lottery.

I would now like to offer two contemporary examples of the application of science to the discovery of new drugs. Excessive production of stomach acid is known to be a factor in the commonly occurring duodenal ulcer, a notoriously recurrent and exceedingly unpleasant condition which can also kill. Recurrent cases can be offered an about 80 per cent chance of cure by surgery. One in 200 patients dies as a result or 10 in 200 if the operation is done as an emergency. So, cure is likely but cannot be guaranteed. But life with the disease can be intolerable or threatened.

The pharmacological approach is to discover the biochemical basis for stomach acid secretion and to make substances that will selectively interfere with this process. This has been done. Cimetidine (Tagamet) relieves ulcer pain and induces healing. When the drug was introduced in 1976, there was an immediate 39 per cent drop in surgical operations for duodenal ulcer. But questions remain to be answered. Does cimetidine eliminate, or merely postpone, the need for surgery? Used continuously, can it prevent indefinitely relapses in this notoriously relapsing condition? Is it safe to use it indefinitely?

The answers cannot be found in the laboratory, but only by studying what happens in large numbers of patients over years. Does this mean that medicines are offered to sick people without knowing what all their effects will be? Yes, it does. Does this mean that large-scale experiment is being done on people with duodenal ulcer? Yes, it does. Is it better to take cimetidine or to have surgery for relapsing duodenal ulcer? We do not know. Would a known death rate from the drug be acceptable? I suspect not, but I cannot tell.

My second example of scientific drug development is that of a new skeletal muscle-relaxing drug, atracurium. The work of the surgeon is eased if our muscles are relaxed, especially in abdominal and thoracic surgery. To achieve this with general anaesthetics means heavy dosage with substantial depression of physiological functions and increased risk of complications and to life itself. How much better to paralyse the patient, to use artificial respiration, and then light drug-induced sleep? But the drugs used to paralyse the patient all have disadvantages, especially in people with poor kidney, liver or circulatory function.

Collaboration

How much better it would be, then, if there were a drug that was eliminated at a predictable rate. Collaboration between the University of Strathclyde and the Wellcome Foundation Limited has resulted in just that. Under the mildly alkaline conditions of the body and assisted by body temperature, the atracurium molecule splits into inert substances so that its action is terminated. By skilful science and molecular manipulation, a substance has been chosen that has the right duration of effect with rapid recovery. No physiological process of the body is required; drug action is the same whether the patient's liver, kidneys and circulation are normal or diseased or merely aged. The invention is brilliant. atracurium exemplifies academic-industrial collaboration at its best.

My final topic is a scientific advance of a different kind. It began in the eighteenth century, was developed in the late nineteenth century and flowered under the impetus of Sir Austin Bradford Hill in the 1940s. It is the accurate measurement of the

efficacy and safety of medicines. If we cannot put numbers to the benefits and to the risks of medicine we cannot make rational choices.

Many medical people, Bradford Hill points out, felt "that there must be some necessary antagonism between the clinical assessment of a few cases and the 'cold mathematics' of the statistically analysed clinical trial dealing with a larger number". But doctors who rely on clinical impressions are also experimenting on their patients, but in an unreliable manner.

Bradford Hill taught us to understand the value of the controlled clinical trial. It is a device of use when we truly do not know whether treatment A is better or worse than

treatment B. I have met people who find it hard to believe doctors when they say that they really do not know. But doctors frequently do *not* know.

What of the next 150 years? For disease for which we currently have medicines I see better medicines. For diseases at present untreatable I see new medicines. But I do not see the elimination of hazard, though I do see better methods of detecting and so of minimizing it. So just as most of us recognize the dangers of motor cars and aeroplanes and gas and electricity in the home but use them thankfully, we should also applaud drugs we now use to attain the Benthamite objective of the greatest happiness for the greatest number. □

Zoological ideas 1831–1981

A.J. Cain

ALTHOUGH it would be an injustice to some great men (Linnaeus, Bonnet, Trembley, Spallanzani, Buffon, Harvey and Haller, for example) to say that there was no modern biology before 1831, it would not be a great injustice. And since one could not survey even the present state of the subject, how much less could one give a history of its development. So I shall look at some of the leading ideas and the major developments, using the state of the science as seen in about 1831, 1881, and 1931.

There is an obvious difference between the inorganic and the natural historical sciences in 1831. There was already an approach to coherent bodies of theory in the inorganic sciences; experiments could be devised to answer particular questions. It has become continually so much more obvious ever since that philosophers often take physics as the paradigm of all sciences. It was simply not true in biology at that time. There was no cell theory, no adequate formulation of homology and analogy, although Cuvier's theory of types was enthusiastically hailed by Dean Buckland at the Oxford meeting in 1832, in words which sound rather odd today.

It was the genius of Cuvier that first established the perfect method after which every succeeding naturalist will model his researches; and which laid the foundation of that analytical process of investigation, of that most philosophical and accurate and uniform system of reducing every organ in every species to a fixed and certain type, which will enable his followers to extend their enquiries over the almost boundless regions of the organized world.

The contemplation of the simplicity of the types, and the real unity within diversity which they reveal, inevitably led Buckland to the conclusion that they were framed and fashioned by the same Almighty hand.

There was, of course, a theory of evolution, Lamarck's, but it was largely disregarded, for scientific as well as religious reasons. There was also, newly available, that most important theory of the uniformity of nature by which Charles

Lyell was refashioning the science of geology; but the gulf between living things and the physical world was so vast that it took many years for its significance to be seen in biology.

But if the plan of creation were of divine origin, there must be some thorough-going order amidst all this bewildering diversity of living things. Louis Agassiz had his own special ideas on this subject, relating the major groups to major developments of the animal functions (British Association, Dublin, 1835). The entomologist W.S. Macleay, on the other hand, had discovered that the animal kingdom was based on the number 5: there were 5 species to each genus, 5 genera to each family, and so on; moreover in each genus, one species was typical and two departed from it in the directions of the two neighbouring genera. Others thought that the number 7 fitted the facts better, and a mycologist plumped for 10.

Let this be an object-lesson to us all never to laugh at outmoded ideas; these exercises completely failed to achieve their ends, but in the course of them the distinction between what Macleay called *affinity* and *analogy* was realized and refined by Richard Owen (1840) into the distinction between *homology* and *analogy*. Moreover, in Hennig's system which is sweeping the museums of the world at the present day, all evolution is dichotomous. (After all, it is the number 2).

On development, there was a conflict of theories. S.D. Broughton at the second meeting of the British Association, came down firmly on the side of the epigeneticists when discussing the development of the ovum; he also showed that muscle fibres were continuous fibres, not rows of globules as some microscopes made them out to be. But there were no great synthetic ideas. On the physiological side, there could hardly be.

Looking over the literature of this time, one is impressed with the intellectual capacity and energy of the zoologists, physiologists and palaeontologists — they were no more fools than we are (and

probably no less). What limited them? It was certainly not lack of ideas nor a neglect of phenomena. Preconceived ideas were important blocks — the idea, from Aristotelian logic that individual variation was accidental (in the logical sense) and therefore had nothing to do with the essence of a species (in the logical and biological senses) prevented most educated people from taking individual variation as anything other than a nuisance. Unfortunately, historians of science who have trained as historians or philosophers are not capable of assessing the limitations imposed by non-existent techniques.

Yet, probably, the biggest single hindrance to the biological sciences in 1831 was the lack of techniques. Biology is peculiarly a subject in which magnification is important in revealing and checking complexity of structure and function. But the extraction, purification and identification of the numerous types of molecules composing the living body is equally important, as is the development of experimental techniques which allow the living animal to be investigated. All this could advance only with the necessary developments in physics, chemistry and technology.

Sheer discovery sustained the developments of the next fifty years. Whereas in physics (perhaps not in astronomy), the role of simple discovery unguided by theory has shrunk almost to nothing, it plays even now a noticeable part in the natural history sciences. Discovery is too intractable for philosophers of science to make anything of, and it has been undervalued in consequence.

After Darwin

By 1881 the zoological scene had changed enormously, but not quite out of all recognition. The exploration of the living and fossil worlds was going on as it is today, but Darwin had published *On the Origin of Species*, Wallace his *Geographical Distribution* and there was a flood of works in physiology, reproduction, and embryology. Microbiology was a science with practical applications, the cell theory was established, and preformation in its older sense had gone the way of spontaneous generation — a few of the older controversies were actually settled.

What the nineteenth century did not bring forward, however, was almost as striking as what it did. Although there was a vast amount of work on plant and animal breeding, the science of genetics did not emerge and for good reasons. In plant breeding the role of insects in fertilizing flowers had not at first been realized, nor the complications in some species of apomixis (with no true crossing) and in others of crossing diploids with high polyploids, giving a "preponderance" of maternal or paternal characters; in animals most of the characters employed were highly polygenic and often susceptible to environmental influence.

Such contradictory results were obtained that no-one knew what to believe. It is not in the least surprising that Mendel's work was overlooked; it is surprising that the value of it was realized as early as 1900.

On the Origin of Species came between the 1831 and 1881 meetings; between 1881 and 1931 we have such a mass of developments that no one work can possibly stand for them by any sort of metonymy, nor indeed would one expect such a phenomenon; in this century big books are common, great books very rare, and advances are made in the scientific journals. Apart from the massive developments in physiology, such as the discovery of hormones and vitamins, there is that sudden outburst of superb experimental work, mainly Belgian, in embryology which produced the fate maps of Vogt, and the concepts of evocation and induction; blood groups begin roughly with Landsteiner in 1900 and immunology starts on its meteoric course; X-ray diffraction is applied to biological materials. Evolution went into the doldrums; in 1909, the half-centenary of the *Origin*, many people were prepared to accept that it had occurred but few would believe in natural selection. Comparative anatomy was, if not worked out, at least no longer exciting the best minds. Ecology was just beginning. The two greatest developments were almost the whole science of genetics, and statistics as applied to biological problems. The old philosophical stumbling-block of induction was at last removed. It is characteristic of the mode of advancement of science that the rediscovery of Mendel's work provoked a strong reaction against Darwin's minute variations and it took twenty years and a rather too vigorous controversy to reconcile them.

Centenary

At the centenary meeting in London (1931), the president, J.C. Smuts, gave an address on "The scientific world picture of today" in which he took the new indeterminacy and relativity of physics, converted them into a philosophy of the wholeness of the organism and assimilated the whole Universe, living and non-living, to it. In doing so he explicitly rejected the nineteenth century determinism of his great predecessor Tyndall's famous address to the British Association in Belfast (1874). Both these addresses show the baleful effect of undigested physics on biology, but its influence goes back much farther than the eighteenth century. Newton remarked that if the complexities of the heavens could be governed by a single short equation, the actual basis of nature might be simpler than anyone had realized.

In Section D at York (1932) the then Lord Rothschild gave a good account of work on geographical variation and speciation. Speciation was a problem untouched by the *Origin* and was one of the main ingredients in the neo-Darwinian

synthesis of the 1930s and 1940s. The discussions in 1931 in Section D included evolution, vertebrate embryology, population, classification with reference to phylogeny and convergence, variation and genetics, the past and present of the over-fishing problem and insects and human welfare. Huxley had been scathing about irresponsible phylogenizing in 1881, just as Medawar was in 1951 in Heath's *Scientific Thought in the Twentieth Century*.

From 1931 to the present day, I hardly need to catalogue the major advances. We have now as great a uniformity throughout the living world in intracellular metabolism as Darwin produced for whole animals by the concept of evolution. But perhaps the most spectacular conversion of a subject from anecdotes to science is in animal behaviour (ethology).

Species

The species problem has been broken up into its constituent parts — there are several sorts of species, and several modes of speciation; the attempt to find a single solution to heterogeneous phenomena is always disastrous in science. And the old problem of spontaneous generation, dead and done with in its Pasteur-Pouchet form, has reappeared in a totally different context, the origin of life, with a real prospect of experimental investigation. Geographical distribution has been revolutionized by plate tectonics, and by the mathematization introduced by MacArthur, Wilson and others.

If there is a problem left over from 1831 and still unsolved, it is indeed the recognition of "true affinity", now usually called phylogenetics. The latest scheme, by Hennig, is constructed entirely from the comparison of characters. The amount of difference between forms is directly proportional to the age of their common ancestor (the further apart, the earlier the ancestor) as if rates of evolution were constant. All evolution is dichotomous, so that if for any group one can identify its sister group, then their common ancestor can be constructed according to set rule. And, most remarkable of all, convergence has not occurred unless it can be shown to have occurred; this is said to be a heuristic principle, without which the basis would be removed from phylogenetic systematics.

Now convergence is being found in all sorts of examples, even where we cannot say exactly which characters have converged. And it is forms already closely related which are most likely to converge, yet in them it is most difficult to tell which are shared ancestral characters and which are convergent. The only thing Hennig has said that is right is that speciation is dichotomous, simply because however many populations of a species are progressing towards species status, it is formally inconceivable that any two should reach it simultaneously; but if several do within the space of a hundred years, one could never detect their order of doing so

on the usually available taxonomic evidence. Moreover, simple dichotomy cannot apply to higher groups, to which it is usually applied. Yet this ridiculous scheme is now being used to reorganize exhibits in museums around the world — a striking example of unreason in science.

Unreason is rife enough in the world outside science. In the past twenty years we have seen a remarkable recrudescence of religious antipathy to evolution. From a different religion have come some intense efforts to prove that natural selection has little importance in evolution. A classic example was the denial on the basis of *a priori* principles, simplistic mathematics and a total refusal actually to investigate the question experimentally, that isozyme polymorphisms, so widespread in living things, could possibly be maintained by selection or be in any sense adaptive.

In fact, those that have been investigated, for example the *Drosophila* alcohol dehydrogenases investigated by Nottingham geneticists, are heavily selected for their different properties in different environments. Lewontin, an ardent supporter of non-adaptive evolution, has endeavoured to dismiss as British upper middle class activity the work of field investigators in this country, which has produced results in contradiction to his tenets. The only question of importance is whether they stand up as science.

Too slow

In Darwin's time no-one worked on natural selection in the field because it was thought that evolution was so slow, it could not be detected. In this century few work on it because to get grants for research, results must be produced within two or three years, and this is long-term work of uncertain outcome. If universities are reduced to poverty, and forced to get all money for research from the government, then such work will become as impossible in Britain as it is in the United States, and the major unanswered question in evolutionary studies — the importance of natural selection in the wild — will continue to be answered by politico-religious dogmatism.

We are closer to a total scientific synthesis in biology than the human race has ever been. No part of it is perfect, and we still wait for major developments in causal embryology and the science of the mind. Yet this is an age of fanaticism, terrorism, fundamentalism. Science has progressed because it depends more on the activities of the gifted few and the practical results which flow from them, in short on cultural evolution which can be very rapid in favourable areas. But the human race has not progressed evolutionarily in the past few hundred years; physical evolution is slow. Are we now at the breaking point between the two evolutions, and will the human race reject its own assessment of the world, to live in a comfortable fantasy? I have no gift of prophecy, and if I had I wouldn't believe it. □

Vasculum to electron microscope

F.R. Whatley

In the mid-nineteenth century, the amateur naturalist and botanist was important. Botanical expeditions, often arranged by local botanical societies, made major scientific contributions to the study of plants and their distribution. They had social as well as scientific value. A curiously significant part of the equipment was the vasculum¹, a tin box carried on a shoulder strap and used for collecting plant specimens that were subsequently identified and annotated, often with great artistry. The vasculum also doubled as a convenient lunch-box.

The importance of the amateur botanist has decreased over the years, but the continuing involvement of the amateur in compiling local floras is significant. Nowadays, scientific plant collecting is often only part of a more highly organized and complicated professional enterprise.

In the fifty years after the founding of the British Association, microscopes in Britain reached a peak of perfection with the introduction of the compound microscope with fully corrected lenses. Cheap microscopes became readily available and came to be regarded as desirable curiosities. However, little progress was made beyond the initial observations made earlier with the simple microscope². Further development of the microscope became centred in Germany, and great advances in the knowledge of plant (and animal) structure as seen in the light microscope took place in a few German universities in the second half of the nineteenth century. The idea was developed that form and function could be correlated; there was also a tradition for teams of scientists to work together in Germany. The resulting encyclopaedic volumes written by Haberlandt³ and others in the 1890s still represent a fount of eminently useful knowledge. The conventional light microscope reached its limits by 1900, by which time the nucleus, mitochondria, chloroplasts and chromosomes had all been recognized as had the behaviour if not the significance of the chromosomes in mitosis and meiosis.

Observations on living material are limited by the degree of contrast. The advent of chemical dyes made the staining of tissues possible. The development of the Mendelian hypothesis and the theory of genes led to a search with stained material for the carriers of the genes and their identification as the chromosomes. The introduction of the fluorescence microscope^{2,4} has allowed other organelles to be identified and characterized. The phase contrast microscope when properly set up allows examination of living material. However, the light microscope is limited by the wavelength of light. Within the past 40 years this limitation has been overcome by the development of the

transmission electron microscope, which uses a beam of electrons at 50–100k V (refs 2,4).

The use of the electron microscope as a tool in biology depended on new methods of fixing tissue, the selection of suitable heavy metal stains to give adequate contrast and the development of embedding materials and of knives capable of cutting ultra-thin sections. The internal fine structure of many of the organelles previously recognized by light microscopy has been revealed and the nature of their membrane systems made clear. Microtubules and previously unknown membranes and other structures have also been identified. The scanning electron microscope is of particular importance in the investigation of surfaces and, of taxonomic significance, "large" objects like pollen grains. The development of the transmission electron microscope to extremely high voltage (1,000kV) offers a new technique for the examination of "thick" sections (1–5 μm) which, when observed as stereoscopic pairs, can give a magnificent indication of the three-dimensional distribution of organelles.

Developments in plant physiology have also depended on the availability and application of new techniques, as illustrated by studies of photosynthesis. Before 1831, the basic features of this fundamental process had been established by the work of Priestley, Ingen-Housz, Senebier and others. There then followed a long hiatus until the application of interrupted light sources by Blackman shortly after 1900 allowed the identification of light and dark reactions and set the scene for the further analysis. Biochemical work established the

evolution of oxygen by isolated chloroplasts (the Hill reaction), the path of carbon (the Calvin cycle — which depended on the availability of $^{14}\text{CO}_2$), the features of photosynthetic phosphorylation (in Arnon's laboratory) and the existence of two light reactions. Current work is largely concerned with control mechanisms in photosynthesis, much of it of physiological importance to agricultural productivity.

Darwin's voyage on the *Beagle* in 1831, the implications of which are still causing reverberations in the literature, was in the tradition of the earlier explorers such as von Humboldt, who were mainly concerned with describing the marvellous new species, plant and animal, to be found in lower latitudes. The monumental work of Schimper⁵, based on observations of many expeditions made in the nineteenth century, correlated the distribution of plants with the habitats in which they are found and stimulated interest in more localized plant distribution within the temperate zone. In the early 1930s, Clements in the United States and, later, Tansley in Britain were among the earliest proponents of the new science of ecology.

With increased awareness of the interdependence of plants and animals and their habitats there has been an increase in public concern about species conservation and the dangers of pollution. It is these aspects of botany which for today's amateur seem to have replaced the Victorian preoccupation with collecting. □

1. Allen, D.E. *Proc. bot. Soc. Br. Is.* 3, 135–150 (1959); 6, 105–109 (1965).
2. Heslop-Harrison, J. *Proc. Roy. Microscop. Soc.* 14, 256–266 (1979).
3. Haberlandt, G. *Physiological Plant Anatomy* (trans. Drummond, M.) (Macmillan, London, 1914).
4. Bradbury, S. *The Evolution of the Microscope* (Pergamon, Oxford, 1967).
5. Schimper, A.F.W. *Plant Geography upon a Physiological Basis* (trans. Fisher, W.R.) (Clarendon, Oxford, 1903).

Sociology — another 150 years?

Philip Abrams

SOCIOLOGY springs from the wish to do something about the dreadful state of the social world — even if only to understand it. Its special characteristic is the hope of rooting whatever it does in knowledge sufficiently cogent, disciplined and objective to pass as science. What distinguishes sociologists from other socially-concerned citizens is the attempt to achieve some demonstration of the science-like nature of their statements about the social world. Unfortunately, that distinction does not save them, as it saves those who wish to make science-like statements about microbes, or planets or fish, from also being part-and-parcel of their own object of study. The problem for sociology is not that of being a science like other sciences but of being a science *appropriate* to the peculiar nature of the social and of sociology's own place in it.

As it happens, 1831 was quite a significant year for sociology; the problem of the relationship between social knowledge and social action was first clearly seen as a problem and defined, albeit mistakenly, as a problem of science. The reasoned analysis of carefully gathered and ordered empirical evidence about society was already widely practised. Such work was something that might be expected of anyone who wanted to contribute seriously to the discussion of matters of state or the condition of society. There was in the 1820s a certain unity of inquiry, analysis and government. By 1831, however, as politics became more open, pluralistic and contested it was a crumbling unity. Social inquiry and analysis came to be seen as controversial and problematic; their relationship to government, and to purposive social action in general, could no

longer be taken for granted; facts ceased to speak for themselves; their very status as facts was called into question.

By 1831, the idea of social science had also appeared but in the form of a radically empiricist activity. Social inquiry was recognized as scientifically legitimate only to the extent that it involved the accumulation of facts. Social science was a passive resource for government, something that “neither discussed causes nor reasoned upon probable effects but sought only to collect, arrange and compare” facts. It was on that basis that Professor Whewell condescended to admit statisticians to the British Association.

By 1881, that conception of a tidy division of labour between fact and action, social scientist and statesman, had collapsed. The search for facts — for example, about the drinking habits of the working classes — had made the world more complicated and all courses of action more controversial. The empiricist idea of social science began to fragment. Determined empiricists turned towards better methods of recording facts, mainly towards refinements of statistical technique. But other voices called for a more overtly theoretical social science.

By 1881, the appetite for a general theory-based science of society had found vigorous expression: it had even been defined as sociology. Thus, at a time when eminent natural scientists were seeking to evict economics and statistics from the British Association on the ground that they were manifestly not sciences, we find the president of Section F urging his peers to turn to sociology to make good the scientific defects of their own work. They ignored his advice, and in one important respect they were right. The prospect J. K. Ingram held out to them of sociology as a positive science reproducing the logic and method of the natural sciences was to prove a calamitous blind alley.

Hypnotic attraction

A hundred years after the publication of *The Wealth of Nations*, economists had failed to convince the scientific community of their own scientific credentials. But the fact that sociology entered intellectual and public history not just in the guise of science but as a science expected to be peculiarly capable of restoring the lost unity of inquiry, analysis and action was thoroughly unfortunate. Yet the invitation to try to be that sort of science was hypnotically attractive.

In 1906, sociology had begun to take on a definite form and content. There was a lively Sociological Society meeting in London, a successful *Sociological Review* and steps had been taken to establish two Professors of Sociology and a Department of Sociology at the London School of Economics. Unfortunately, institutional development seemed already to be marching hand-in-hand with intellectual disarray.

The nature and province of the emergent discipline were the heart of the problem. That a social science was desirable was widely agreed. But just what would be involved? Positivism, empiricism, the ameliorist urge to do something quickly about pressing social problems and sheer vested interests — already established academic disciplines — combined to discourage sensible answers to such questions. Prevailing empiricist conceptions of science, the appetite for definitive information about poverty, industrial conflict, town planning and a host of other current social problems, and the mere practicalities of academic politics persuaded many to see sociology as little more than a battery of research techniques occupying a province of substantive problems which had a purely *ad hoc* existence. The social came to be defined as a miscellaneous grouping-together of whatever economics, history and psychology between them had not already appropriated. Trapped thus, between contrary expectations and possibilities, sociology was by 1906 already busy sowing the seeds of its own subsequent crises.

By 1906, there were as many divergent definitions of sociology as there were sociologists. Already one could hear loud complaints about the uselessness of this variety of sociology, the arid pedantry of that, the misty philosophizing of another, the political tendentiousness of yet another. Indeed, the disputes of the 1970s seem but genteel hiccups compared with the fratricidal strife of sociologists of the Edwardian period. For if sociology was not a general science of the social, what on earth was it? In the event the price of scientific respectability was paid. Sociology came to refer to an absurdly delimited province of inquiry investigated on the basis of absurd expectations about the possibility, in social inquiry, of access to both laws and facts.

In 1956, the situation was superficially much the same. As a discipline, sociology still had only a marginal institutional existence. Ten universities made some provision for the teaching of sociology. Intellectually the divisive tendencies among sociologists were as strong as ever. Although T. H. Marshall had diagnosed the problem acutely in his Inaugural Lecture at the London School of Economics in 1946, calling on sociologists to abandon both “the (positivist) ways to the stars . . . vast generalizations, universal laws” and the empiricist “way into the sands of whirling facts which blow into the eyes until nothing can be clearly seen”, and to follow instead a middle way of “manageable units of study”, “specific social structures” and “determined meanings”, there was hardly a rush to take his advice.

Looking back in 1962, W. J. H. Sprott saw sociologists in the intervening years not at Marshall’s “crossroads . . . a body of men marching together and faced with a

decision”, but rather as a crowd uncertainly milling to and fro at the Seven Dials . . . “several bodies of men converging on an open space where they spend a good deal of time abusing one another”. Few would have thought in 1956 that a great intellectual and institutional explosion of sociology was imminent.

Sociology established

Yet by 1956 there had been far-reaching changes in the social standing of sociology. Two developments were particularly important: in the relatively purposeful political context that existed in Britain from 1940 to 1950, sociologists had managed to demonstrate that their work could have a certain usefulness; and then, since 1950, a marked loss of political purpose, a growing uncertainty of values and a general confusion about the nature of society had given a distinct edge and urgency to questions of a distinctly sociological nature. Amid the confusion at the Seven Dials, both positivism and empiricism had been held at bay, at least enough to allow some middle ground to be cultivated. It seems to have been very much in that spirit that Lord Adrian, as President of the British Association in 1954, urged “full contact” between the social sciences and what he called “all the conservative and academic people in universities”.

Fifteen years ago, we find an amazing transformation; sociology was fully established in our universities. Many of the complaints about the turbulence, incoherence, mystification and uselessness of the subject lose their force when one remembers that sociology, as a nationally established discipline, is still in its first twenty years. Marshall’s lead notwithstanding, there was an enormous amount of intellectual ground-clearing still to be done; the credibility of all sorts of competing accounts of sociology had still to be explored; the ghosts of positivism and empiricism had still to be laid; the sensible limits of usefulness had to be clarified; and until 1960 there were simply not enough people to do the work. The question we can now ask, however, is whether that short but intense period of self-proclaimed crisis is to be understood as the natural condition of sociology, revealing the impossibility of its claims to scientific status, or as something else.

The question is important because public enthusiasm for sociology is now on the wane. Whether the history of sociology — in particular the brief tumultuous history of the past 20 years — reveals the absurdity of its scientific pretensions is crucially important. It is the question on which the future of the discipline seems to turn.

There are perhaps two separate questions. One is about sociology as a diffuse practice of social inquiry and analysis. The other is about the existence of sociology as a special, fairly narrowly defined and academic discipline of thought. We could decide that there are

good reasons for keeping both.

For most of the past 150 years sociologists have had to struggle against prevailing conceptions of science and of society and above all against prevailing expectations about a possible science of society. Called into being to be something it could never be, sociology has had to spend every period of encouragement it has been given re-discovering the insights of the eighteenth century. Successive moments of disillusionment with sociology are also the moments at which the performance of sociology has managed to give the lie to social expectations of it. But, of course, to be disillusioned is also a form of scientific progress. The real object of criticism should be the illusory expectations.

History of crises

It is my view that the history of sociology is indeed a history of crises of disillusionment, and yet those very crises have been both its necessary form of existence and the way in which it has achieved scientific growth. What we can now see in this history is a progressive discovery of both the possibility and the limits of social knowledge. Each moment of encouragement — 1881, 1906, 1945, 1960 — saw an explosive attack on the possibility of providing types of authoritative knowledge that were widely hoped for but not in fact attainable.

The army of young scholars and enthusiasts who entered the discipline after 1960 have cleared away, at last, a great deal of the intellectual lumber of the past. The rival factions and schools of thought have managed in their furious and perceptive attacks on one another to put paid to a lot of nonsense. Almost all the possible ways of saving positivism, empiricism and authoritative knowledge of the social have surely now been ruthlessly criticized and put out of court. But what we are left with is not uselessness but a more restricted usefulness, the usefulness of informed scepticism and reasoned advocacy.

What Michael Mann has called a common “sociologic” has emerged as constituting for all schools a proper, self-conscious, self-monitoring, rigorous and effective procedure of inquiry and analysis. Simultaneously, as one would expect, knowledge itself has accumulated. And again it is knowledge adapted both to the nature of the social as an object of analysis and to the limits of sociology as a form of analysis set within its own object of inquiry.

Sociologists manifestly do know more than politicians about the causes of urban riots, even though their knowledge does not come in the form of a recipe for riot prevention. They have vastly deepened and refined our possible understanding of social inequality even though they have left open every question about what to do about inequality. They have helped to make a whole series of political myths untenable. John Rex has described

sociology as a work of demystification. There is more to it than that, but even in achieving so much demystification in so short a time, sociology is well up to the track record of, say, physics or chemistry or geology.

But sociology has also discovered that in the social sciences, knowledge is *not* power. Sociological knowledge is relevant to policy and is deployed within the arena of politics. But it cannot hope to be either politically neutral or politically authoritative. The performance of sociology since 1960 has dashed many of the expectations of that period irrevocably but constructively. It remains to be seen whether the ground-clearing sociologists thus achieved will be recognized by others. It remains, that is to say, for others to lose their illusions about sociology just as sociologists have done. Government longs for certainty; but in social analysis certainty is not to be had; sociology is an active resource for politics, not a passive

resource for government.

If the achievement of the past 20 years of sustained large-scale professional work in sociology had been no more than to purge the errors of the past and place our ideas of the nature and province of sociology on a sensibly scientific basis, it would have been considerable. Behind the facade of crisis that sort of reconstitution of the field both as to method and in terms of substantive work certainly has taken place. But more has been done. These have not just been years of debates about principles. The first 20 years of sensible sociology in Britain have also seen, thanks to the new scale and concentration of intellectual resources within the discipline, the production of a great deal of detailed social inquiry and analysis. Sociology is now in my view ready, both methodologically and substantively, to perform as a reasonable science of society. What matters is that expectations should be brought into line with that possible performance. □

Men, bishops and apes

Edmund R. Leach

IN this symposium I am representing anthropology considered as a social science. That is indeed how most contemporary anthropologists would like to regard their subject but members of the general public mostly have a different view. Insofar as there are misunderstandings on that point, the British Association is partly to blame. That circumstance explains my title and will provide my focus.

Although it was the notorious confrontation between T.H. Huxley and Bishop Wilberforce at the British Association meeting at Oxford in 1860 which first put anthropology onto the academic map, anthropologists have not been prominent in the affairs of the association. Their most notable contributions were at Montreal in 1884, which saw the creation of the present Section H under the presidency of E. B. Tylor, and at the meeting of 1914 held in Australia when the participants included W. H. R. Rivers, Bronislaw Malinowski and A. R. Radcliffe-Brown, then known as plain Mr Brown.

The printed records of the British Association tell us little. The addresses by the presidents of Section H were usually printed in full but other papers appear only under their bare titles or as an abstract. If a paper provoked discussion, there is no record of this fact. And indeed the Huxley/Wilberforce confrontation of 1860, though doubtless carefully planned, developed as an argument from the floor about two separate papers which were quite uninteresting in themselves. Consequently the records that we have of that event have been mythological from the start.

The myth appears in various versions but deserves repetition. Darwin's *The Origin*

of *Species* was published in 1859. Darwin was already recognized as the most distinguished naturalist of his day. The essentials of the Darwin-Wallace theory of natural selection had been published two years before but the appearance of the full-scale book immediately created a sensation. Although Darwin himself here evaded the issue of the origin of man, the theological significance of his theory was immediately recognized. By implication, man was a near-ape modified by natural selection; this was palpable heresy, totally incompatible with the dogma that man is a special creation designed from scratch “in the image of God”.

Bishop Wilberforce's review of Darwin's book in *Quarterly Review* had been written before the Oxford meeting of the British Association but was published afterwards. It is a much more impressive document than recent abstracts suggest. Scientists of the highest standing would today accept many of Wilberforce's criticisms of Darwin just as they would also accept the criticisms raised by the geologist Adam Sedgwick, whose review was published in *The Spectator* in April 1860.

Indeed I suspect that had Darwin known as much as we do about the mechanism of genetic mutations and about the recurrent pattern in the fossil record of long-term stability followed by sudden change, he himself, had he been alive, would have been a convert to the neo-Sedgwickian views now being advanced from Harvard by Stephen Jay Gould. For despite his heresies Gould is not essentially a Darwinian. The point is simply that orthodox Darwinian theory maintains that evolution occurred as a gradual process of innumerable very small changes. Darwin's original

opponents stood for sudden change, either as a consequence of divine intervention or as the result of natural catastrophe. But Darwin adapted his gradualist view of evolution mainly because he could not imagine how any biological mechanism could create the appearance of sudden discontinuity and he emphatically did not want to believe in divine intervention.

However, at the time, the issue was seen by Darwin's defenders, such as T.H. Huxley, as a straight conflict between bigotry and enlightened rationalism. Darwin's theory was misinterpreted as a vindication of the political doctrine that a policy of *laissez-faire* will inevitably result in Progress with a capital 'P': on and on and on and up and up and up.

What Darwin's theory really said was that species tend to become adapted to increasingly specialized environments; it was not concerned with the notion of "Progress" at all. Yet one still encounters biologists who refer to "advanced" and "primitive" species and who are quite evidently predisposed to think that mankind can be sub-divided into "advanced" and "primitive" "races", "cultures" or what have you. Needless to say it is "we" who are "advanced", "we" being the particular human collectivity to which the author himself happens to belong.

The difficulty about applying Darwin's theory to man in any simple way should have been obvious from the start since our human peculiarity is that we are remarkably unspecialized. Even without the aid of artificial gadgetry, human beings are able to survive and flourish in forests and in deserts, in the tropics and in the Arctic. Apart from creatures such as the rat and the dog, which seem to have evolved as parasites of man, no other natural species has this kind of flexibility. So if man evolved from his predecessors by natural selection associated with adaptation to a specialized environment, what kind of environment was that?

Most of what has been written on this topic in recent years seems to me quite astonishingly naive. If the first men were well adapted to living in African savannahs, how did their descendants manage to flourish among Greenland's icy mountains or in Europe's ice ages?

But to get back to the Wilberforce/Huxley affair of 1860. In most respects the debate has now been turned back to front. The evolution of species from earlier species is not seriously questioned; nor is the theory that most species are specially adapted to the environmental niche in which they are encountered. But it is becoming increasingly difficult to understand just how they came to be that way. "Evolution by natural selection and adaptation" seems to be a much more complex and variable process than had at one time been supposed.

In particular, Darwin's central doctrine of development by numberless minute changes has been coming under sharp attack. It now seems that if general evolutionary theory is to

be compatible with the geological and palaeontological evidence it will have to incorporate some kind of modified version of Sedgwick's catastrophism.

Missing links

Missing links in the sequence of fossil evidence were a worry to Darwin. He felt sure they would eventually turn up, but they are still missing and seem likely to remain so. What we are to make of that fact is still open to debate, but today it is the conventional neo-Darwinians who appear as the conservative bigots and the unorthodox neo-Sedgwickians who rate as enlightened rationalists prepared to contemplate the evidence that is plain for all to see.

But to return to the myth of the Wilberforce/Huxley confrontation. The version I like best has Wilberforce proclaim: "Whatever certain people may believe I would not look at the monkeys in the Zoological Gardens as connected with my ancestors"; to which Huxley responded: "I would rather be descended from an ape than a bishop". Which is good as repartee, but hardly a contribution to science.

Indirectly this celebrated episode had regrettable consequences for the academic subject of anthropology. In the first place it fixed in the public mind the idea, which has prevailed to the present day, that the "typical anthropologist" is a fossil hunter like Richard Leakey. In fact fossil hunting and speculation about the origins of mankind is only a tiny, rather off-beat, sub-branch of the general field of anthropology. The main concern of anthropologists is with the variety of human culture rather than the variety of human skulls.

One of the tasks of anthropology is to provide a bridge between the biological and the social sciences, but for a long time the great majority of anthropologists have thought of themselves as concerned with problems of human society rather than with peculiarities of our anatomy. Furthermore, the nineteenth century model in which society was thought of as a kind of living organism, which had evolved according to Darwinian principles, has long been abandoned, though Marxist versions of social Darwinism are still with us. But even in the late nineteenth century, as Tylor's Montreal address makes plain, British anthropology was primarily a social rather than a biological science.

There have been major changes in objective. Tylor and his contemporaries imagined that by studying the customs of primitive peoples the anthropologist would be able to reconstruct the sequences of pre-history much as the early Darwinians thought they could reconstruct the sequences of species development. That is why Section H of the British Association is now primarily a forum for local archaeologists. Thus, here at York, the meetings of this section are focused on the activities of the Vikings. By contrast professional anthropologists would now

ordinarily concern themselves with aspects of present day culture. The latter are interested in historical evidence for comparative purposes but they no longer claim to be engaged in a kind of archaeology of custom. But the general public still imagines that anthropology is more concerned with human origins than with comparative sociology.

Another regrettable aspect of the 1860 confrontation was that it created a fixed illusion that all anthropologists are aggressively polemical anti-religious materialists. In fact, at a personal level, the religious views of anthropologists are as varied as those of any other section of the community. But the phenomena which get slotted into the category "religion" include many of the most distinctive features of human culture. On that account all academic anthropologists pay close attention to the details of religious practice and belief. Far from being generally "hostile" to religion, anthropologists are almost the only academics, other than divinity school theologians, who take the study of religion seriously.

What is science?

All this has wide relevance for the present occasion and not only for the anthropologists. During most of the last 150 years it has been widely believed that all professional scientists (including of course those associated with the British Association) are wholly committed to a godless materialism. In this context Huxley's verbal encounter with Wilberforce has repeatedly been used as a symbol of the triumph of scientific honesty over the obscurantism of religious prejudice.

But where do we stand today? If the humanist anthropologists share the theologian's interest in comparative religion, does that mean that they must now be at war with the pure and applied scientists? And does that, in turn, imply that anthropology no longer has any proper place here in the British Association? Or, to put it rather differently: in the context of the 150th anniversary of the British Association for the Advancement of Science what does the word "science" really mean?

The issue is a particularly sensitive one for the various social sciences. Strictly speaking, "science" simply means "knowledge", and any branch of learning which is scrupulous about the way it interprets evidence can properly be rated "scientific". But, in England (and indeed in most of Europe), a major distinction is drawn between the natural sciences on the one hand and the humanities on the other. In this great divide, all the social sciences represented in this symposium are now rated among the humanities although not so very long ago the anthropologists (minus the archaeologists) were classed as natural scientists. This ambiguity deserves attention.

Ideology

In my own view the social scientists, so called, have been most ill advised in the way they have attempted to imitate, not only the methods, but the whole ideology of the natural sciences. That ideology has been scarcely modified since the days of Descartes at the beginning of the seventeenth century. "Natural scientists" take it for granted that all the processes of nature are governed by "laws" which are not open to manipulation; it is the general task of scientists to discover what those "laws" are. In effect, the scientific observer perceives himself as an entirely neutral external figure who can experiment upon nature without interacting with it. It is taken for granted that the materials that are subject to such experiments have no will of their own; they cannot tell lies; they cannot alter the rules of the game in the course of the experiment.

The main justification for this belief is that in many fields it seems to work. Relying on the mechanistic certainty of scientific laws we are now able to send men to the Moon and bring them back again with a margin of error of a few seconds and a few hundred yards. If anything goes wrong in such demonstrations of infallibility the defect is attributed to "human error". That is very significant. Nature cannot tell lies but human beings can and do.

And that is the central paradox of the social sciences. At a certain level the economists and the sociologists and even the anthropologists seem to talk as if the behaviour of human individuals and human collectivities was governed by "laws" which are analogous to the second law of thermodynamics. Yet they have never managed to demonstrate that any such laws exist.

The difference in this respect between the natural sciences and the social sciences is profound. Natural science progresses whereas the social scientists go on fighting the same intellectual battles over and over again. The most influential sociologist of the present day is Karl Marx who was already dead when Tylor held the first presidency of our Section H.

I do not find this surprising. The mechanical models of natural science are inappropriate for the social sciences. Human beings engage in wilful deception on a massive scale. The ability to tell lies is perhaps our most striking human characteristic. The dogma invoked in economics, and sometimes elsewhere, that, by resort to statistics and the mathematics of probability, we can eliminate the consequences of this wilfulness and thus reduce human behaviour to law-bound regularity, is just wishful thinking.

Human experience is everywhere chaotic. This chaos is psychologically disturbing. All human culture is designed to make the chaos appear more orderly than it is. Nature-out-there is full of

random shapes and tangled intricacies; most of the material products of human culture employ only the very simplest forms of geometry. Likewise, in interpreting our sensory experiences, we habitually operate with the prime assumption that, whatever the appearances, Nature is "really" constructed in quite a simple fashion according to man-made mechanical principles.

There is no harm in such an assumption as long as it works. And it does usually work as long as creatures with wills of their own are kept out of the picture. But if we try to extend our interpretative observations to the behaviour of human beings, the mechanical model of natural science always breaks down. Human beings are different.

How are they different? The traditional religious answer was that men are different because they were "made in the image of God". This was what the Wilberforce/Huxley debate was about. If Darwin's scheme of gradualist evolution were to be accepted there could be no discontinuity. If men had souls, so had chimpanzees. Such a doctrine would be perfectly acceptable to Hindus and Buddhists but it was more than Wilberforce could stand. And I think Wilberforce was right.

I do not myself believe that human beings have immortal souls but they unquestionably possess spoken language and this, in my view, makes them creatures of a quite unique and novel type. For it is the very special kind of thinking which

derives from our possession of language which allows us to plan deception on a massive scale.

I am aware of course that some primatologists question the uniqueness of human speech, claiming that the circus tricks learnt by Washoe the chimpanzee, Koko the gorilla and their various expensively trained companions represent an embryo form of speech. This is a highly technical matter but, as far as I can judge, almost all the psychologists and linguists who have studied the evidence closely are quite satisfied that there is a major discontinuity between human speech and the sign-making capacities of apes.

Bishop Wilberforce would have put the matter much more simply. He would have said: "If you think that the difference between yourself and Koko the gorilla is simply one of degree, then do you consider that Koko the gorilla could become a Christian?". If Huxley had then replied: "But I do not want to be a Christian anyway", Wilberforce's obvious response would have been: "But you can choose; Koko the gorilla cannot".

And that, at the end of the day, is why the social sciences can never be sciences in the natural science sense. Human beings can make conscious choices. The capacity to make such choices, which is linked with language, represents a major discontinuity with the rest of nature. It has the implication that the only "laws" of social science are matters of the most banal triviality. □

From Great Duke to microchip

Ewen McEwen

ARTHUR, Duke of Wellington, ceased to be Prime Minister in 1830, the year before the foundation of the British Association for the Advancement of Science; he was then 61 years of age and lived until 1852. His departure from office marked the end of an era and the new Victorian age, shortly to begin, in which engineering made a tremendous contribution. The early Victorian engineer was much more of a public hero than his successors have been.

The Duke was not only a good horseman but became very much attached to his particular horses. To him, the horses was the principal means of transport. Yet the reign of the horse even in agriculture was fairly short, as has been the case with other sources of power in other applications: the horse became the dominant source of power on the farm only during the agricultural revolution of the 1730s.

Wellington, however, saw the beginning of the railway era. The Stockton & Darlington Railway opened in 1825, and the Liverpool & Manchester in 1830 both of them built by George Stephenson, whose bicentenary we celebrate this year. Isambard Brunel was appointed Chief

Engineer and began the construction of the Great Western Railway in 1833, when he was 27. The reciprocating steam engine as developed by Watt and improved for railway use by such men as Stephenson and Gooch remained the major prime mover for railway use for more than a century.

It is interesting that an apparently inefficient device should have lasted so long. The cost of fuel in such countries as Britain may have been only a small part of the cost of operating a railway, but this was not true everywhere. Fuel was a major cost in many cases, as for example on the Guaqui-La Paz railway, built by my grandfather, John Pierce-Hope. In the period between the wars, the development of steam turbines seemed to offer a promising alternative, but there was a major snag: a steam turbine locomotive must use air cooling for its condenser and the energy consumption of the fans virtually eliminated the energy saving. It was also necessary to use forced ventilation to produce sufficient draught. Thus the end result was an increase in complexity for no real saving in fuel.

The success of the railways depended not only on their source of motive power but on their right of way. The railway builders cut, tunnelled, bridged, embanked and switchbacked their permanent ways so as to minimize the gradients and enable a locomotive of relatively limited tractive effort to haul "dead" cars and waggons.

The horse, however, continued to dominate road transport. The exertions of Telford in the last years of the eighteenth century and the early 1800s in road and bridge building, improved by Macadam's introduction of better surfacing from 1810 onwards, transformed the roads to all-weather usability and a great improvement in feeder transport to and from the railways resulted. But the horse-drawn coach could negotiate fairly steep gradients so it was not thought necessary or cost effective to build roads to railway standards.

Nevertheless, in the early nineteenth century a number of men (Medhurst (1819), Griffin (1821), Brown (1823) and Gordon (1824)) turned their attention to the possibilities of steam carriages. They failed for a variety of reasons but basically because the steam engine could not be efficiently and practically made small enough. Even the recent Ford experiments have not solved this problem.

It is not surprising that the development of coal gas, primarily for lighting, in the eighteenth century should have prompted interest in its use as a fuel for prime movers. Barker patented the gas turbine in 1791. He was before his time for the necessary materials were not available, but right in his belief that external combustion was necessary — there was no available means of ignition for internal combustion.

Ignition

The internal combustion engine had to await the inventions of Faraday, Ampère and Pixii (1831) to provide spark ignition. In 1860 Lenoir patented a two-stroke gas engine without compression with electric coil ignition. It was the first commercially successful internal combustion engine but still of such a size and weight as to be suitable only for stationary work. Ten years later Hoch made the first practical petroleum engine. Then in 1873 Rothman put into commercial service a four-stroke engine — which had been proposed by the Rochas in 1862 — and three years later Otto began his production of four-stroke gas engines. In 1883 Daimler produced his high-speed (800 r.p.m.) petrol engine, and in 1885 Benz built his tricycle motor car, but the motor car remained a rich man's toy until Maybach (1893) invented the jet carburettor and Robert Bosch (1902) the magneto and the modern spark plug. The way was open for Henry Ford to introduce in 1908 the mass-produced motor car.

The steam engine took more successfully to sea than to the roads. Papin built a steam paddle boat in 1707 but this was an atmospheric engine and very inefficient. The first commercial steam boat, using a

Watt-type engine, was built by Fulton in 1807. Developments in superheating and multiple expansion engines kept the reciprocating steam engine in power as the main method of ship propulsion from then until the early years of the twentieth century, when it began to give way first to the steam turbine, then to the marine diesel and finally, in modern warships, to the gas turbine.

It is interesting that a steam turbine, basically a toy, was built by Hero of Alexandria about 50 AD; that Barber proposed a gas turbine in 1791 and that Trevithick built a reaction steam turbine as early as 1815, the year of Waterloo. Why the long delays? The real reason that the steam turbine could not, in its early days, compete with the reciprocating steam engine was that sufficiently accurate gearing was not available to allow the turbine to run at high enough speeds to obtain adequate efficiency. High-speed helical gearing was introduced by de Laval in 1890.

Gas turbines

The story of the gas turbine hinges on the availability of blade and combustion chamber materials capable of resisting the hot corrosive gases. Modern ceramic materials may well enable the gas turbine to make more inroads into what is currently done by reciprocating engines, by improving the thermal efficiency. What is really needed, however, is an adequate robust reliable heat exchanger to recoup the energy in the exhaust. The development of such a device would have as dramatic an effect on the fuel consumption as did Watt's invention of the steam condenser.

The development of aviation had to await prime movers of adequate power to weight ratio: early efforts to fly were limited by the power available from the pilot quite apart from a lack of appreciation of aerodynamics. The development of the motor car made available the petrol internal combustion engine with a reasonably high power to weight ratio — including, of course, the weight of the fuel. In 1900 Zeppelin used two 16 hp Daimler engines to drive a dirigible airship and in 1903 the Wright brothers used their own 12 hp engine for the first flight of a heavier than air machine.

In agriculture, although the horse continued to be the main source of power, ploughing by cable and steam traction engines had a limited use in the nineteenth century. Fuel and water supply were, however, a serious limitation and generally speaking the available implements were unsuitable for mechanical traction: in 1827, Hawkins remarked that it was just as unreasonable to expect to plough and cultivate by steam as it was to expect horses to dig with spades. Mechanical devices were nevertheless appearing on the farm, but were horsedrawn: the Reverend Patrick Bell invented a reaper in 1827,

Cyrus McCormick the reaper and binder in 1834 and Mackay the header/thresher combine harvester in 1845.

The late Professor Ormsby used to tell his civil engineering students that the civil engineer had done more for public health with the provision of clean water and proper sewerage than all the physicians. The provision of aqueducts and reservoirs is, of course, of considerable antiquity.

The first dams on modern scientific principles were built in France in the latter half of the nineteenth century, beginning with the Zola dam in the 1850s near Aix-en-Provence — an arch dam — and the great gravity dams of which the prototype was Graef and Delocre's famous dam for the water supply of St Etienne (1861–66).

Civil engineers have thus given us water and sewerage and made the railway viable by the construction of the permanent way, which required a vast improvement in bridge building both for economic and topographic reasons. Darby built the first iron bridge over the Severn (an arched bridge) in 1777–79 and Telford spanned the Menai Straits with a suspension bridge built between 1819 and 1826. The earliest important iron girder bridge was Robert Stephenson's Britannia tubular bridge, also over the Menai Straits (1846–50), for the Holyhead railway. But all was not always well. The aerodynamic problems of suspension bridges were long not fully understood: the Union Bridge for the railway at Berwick upon Tweed failed after six months and as recently as 1940 there was the Tacoma Narrows disaster in the United States.

Military angle

Telecommunication has long been a human want much emphasized by military necessity and later by the operation of the railways. The signal fires of the Armada and the semaphore telegraphy of the escape of Napoleon from Elba are among the forerunners. But in 1857, the year of the accession of Queen Victoria, Cooke and Wheatstone combined to bring the electric telegraph into daily use: a few years later Morse established it — and his code — in the Western Hemisphere. It was soon realized that the capacity of the electrical circuit was much greater than that of the human operator, which led Wheatstone in 1858 to the automatic telegraph.

Having achieved electric telegraphy overland, there was an immediate demand to make it available across water. A Commons Committee heard Wheatstone's views on this in 1840; the Brett brothers laid a submarine cable from Dover to Gris Nez in 1850 which quickly failed, but the following year they laid an improved cable which remained in service for many years. The problems of laying a cable under the Atlantic were, of course, more difficult; nevertheless in 1858 the connection was made, although it failed after a few weeks.

"Written" telecommunications were soon followed by verbal and sound

communication: Graham Bell invented the telephone in 1875 and by 1878 the idea of the telephone exchange had been proposed and the first (at New Haven, Connecticut) installed. Strowger invented the electro-mechanical system which bears his name in 1889: the first automatic exchange was installed at Augusta (New York) in 1897 and the first in England at Epsom in 1912. The Strowger system is only just out of use in favour of electronic systems.

In the earlier forms of telecommunication the engineer made inventions which the scientist then analysed, establishing a theoretical foundation on which improvements could be based. Radio is a notable exception: here the theoretical work of Clerk Maxwell (1864) and Hertz (1891) was practically applied by Marconi, who made an effective radio transmission over a distance of one kilometre in 1895. The first commercial use was in 1897 between Poole and the Isle of Wight. The invention of the "valve" — originally a diode rectifier — by Fleming in 1904 and of the triode by de Forest two years later soon produced major improvements in radio communication: by 1910 many passenger ships were equipped with wireless telegraphy. The era of electronic entertainment began in 1920, and in 1922 the British Broadcasting Corporation was formed — to begin with as a limited company.

The radio valve made possible radio and television — and much else besides. But it was large and fragile: aircraft engine management systems were hydromechanical because of the vulnerability and consequent unreliability of electronic systems based on valves. There were many applications where their sheer bulk prevented their use. But then, in 1948 Shockley, Barden and Brattain produced the germanium transistor triode. This was the major breakthrough.

Other semiconductors were tried out and silicon replaced germanium. Methods of miniaturization were developed and multilayer techniques led to integrated circuitry, then LSI (large scale integration) and so to the microchip.

And so we come to the microchip, the miniature computer which is so much more powerful than the early main frame machines. We have opened Pandora's box. We have a tool more powerful than any we have had before and I see no sign anywhere in the world that any nation understands the revolution that now faces us. It seems to me that politicians, economists, employers, unions — and scientists and engineers — must appreciate that it is a completely new era. Minor changes in productivity, facelifts in the product, the common currency of tax changes — all the ways in which we have made, or failed to make, progress in the past are but a sigh in a hurricane. When are we going to apply our minds to the biggest challenge that the human race has ever faced? □

Some consequences of physics

R.V. Jones

ON 13 June 1831 James Clerk Maxwell was born; and on 29 August Michael Faraday discovered electromagnetic induction. If the latter event shows the state of physics when the British Association first met, the former shows its promise.

Newtonian mechanics had long been established; and the inverse square law of gravitation had been found to have its parallels in magnetic and electrostatic attraction and repulsion, enabling simple electrical phenomena to be expressed in mechanical, and therefore in mathematical, terms. In the first decade of the century the corpuscular theory of light had apparently been overthrown by the wave theory of Thomas Young; and the visible spectrum had been extended into the infrared by William Herschel and into the ultraviolet by J.W. Ritter. Joseph Priestley had argued as early as 1767 that there should be connections between electricity and both chemistry and physiology; and the 1830s saw the former prediction borne out by Faraday's laws of electrolysis.

Yet another of Faraday's discoveries of the 1830s was to prove seminal — the electrical discharge in a gas at low pressure, with the colour of its glowing positive column dependent on the chemical nature of the gas. It is possible to trace forward from that discovery through the work of Geissler, Plücker, Hittorf and Crookes to the discovery of cathode rays and positive rays, and hence on to J.J. Thomson and the electron.

In 1831, three conservation laws — those of mass, momentum and mechanical energy — appeared well established; and by 1850 the last of these had been greatly extended by Joule and Mayer to the generalized conservation of energy which put the thermodynamics of Carnot on a firm basis. The second law of thermodynamics was a "postulate of impotence" which later started a train of thought in the mind of Einstein¹, who argued by analogy that a similar restricting principle applied to the Lorentz transformation, with the result that it is impossible for an observer to detect absolute motion through space.

The researches in optics destined to climax in the Michelson–Morley experiments of the 1880s were under way. By analogy with sound, Doppler had in 1842 postulated his effect as the explanation of why some stars appeared reddish and others blue, however Fizeau showed that the Doppler effect could not explain such a gross phenomenon, but that the effect might show up in the sharp lines of stellar spectra. Fizeau, a magnificent experimenter, also detected the Fresnel "aether drag" with an interferometer and, building on this work, Michelson went on to his famous experiment on aether drift.

At the 1859 meeting of the British Association, Maxwell announced his law

governing the displacement of molecular velocities in gases. Besides setting the pattern for subsequent distribution laws and the statistical approach to physical phenomena, this led him to predict — astonishingly but correctly — that the viscosity of a gas should, over a large range, be independent of its pressure, a prediction of which Rayleigh² later wrote that physics contained "no more beautiful or telling discovery" because it established the kinetic theory beyond doubt.

Even this achievement, though, was within a very few years to be eclipsed by the greatest of all advances made in the nineteenth century — again by Maxwell. As early as 1861, he had formulated the framework of the electromagnetic theory which led him to conclude that light was an electromagnetic phenomenon propagated according to electromagnetic laws, and to point out that the aetherial medium might be capable of other motions not appreciable by our senses. It was this prediction that led Hertz to look for oscillatory waves arising from electrical discharges and thus to his discovery of radio waves in 1887.

Television

These two great physicists, Maxwell and Hertz, each made further pregnant discoveries. Maxwell advanced the three-colour theory to the point where he was able to demonstrate the first colour photograph in 1861. And Hertz, in an incidental but brilliant observation of the seemingly irregular behaviour of the spark gap that he used as detector in his radio experiments, discovered the external photoelectric effect. These two discoveries, along with those of radio waves and of the electron, were essential to the invention of colour television with all its social consequences.

Progress in infrared studies was much less spectacular because it was necessarily painstaking. But the result was the black body curve, which defied classical explanation. Matching his experimental colleagues in patience, Planck tried one hypothesis after another until he found in 1900 that the shape of the curve could be explained if radiation were exchanged between a material surface and free space only in discrete packets or quanta of energy. This was the dawn of quantum theory.

Special relativity, too, was about to dawn. Einstein had been deeply impressed with Maxwell's electromagnetic theory, and he opened his famous 1905 paper on relativity by remarking how he had been struck by the asymmetrical explanations of the phenomenon of inducing a current in a circuit with a magnet depending on whether one regards the magnet as moving and the circuit as fixed or vice versa. Einstein later³ wrote "the special theory of

relativity owes its origins to the Maxwell equations of the electromagnetic field"; and one of the consequences that he foresaw was his equation $E = mc^2$.

In what was to become atomic physics, the work of Rowland and Ångström in measuring the wavelengths of structural lines led Balmer in 1885 to his empirical formula for the wavelengths emitted by the hydrogen atom. And although for some time it appeared to be in a quite different field, Röntgen's discovery of X rays in 1895 bore remarkable fruit. Within eighteen months of that discovery the mobile surgical units accompanying the British expedition to Omdurman had X-ray equipment for examining wounds; and it was while investigating the ionization caused by X rays that J.J. Thomson was led to his work on the electron.

Becquerel's discovery of radioactivity in 1896 was equally fruitful, for it led to Rutherford's recognition of the atomic nucleus and thus to the whole of nuclear physics. Aston's skilful development of the mass spectrometer led to his discovery of the nuclear mass defect in the middle of the Periodic Table and thus to Eddington's pronouncement (at the 1920 meeting of the British Association) that this defect could be explained by assuming that enormous energies were involved in nuclear fission and fusion, thus pointing the way to nuclear power and nuclear bombs and to the understanding of mechanisms by which energy is released in the stars.

There followed the recognition by de Broglie in 1924 that, as light had been found to show both wave and particle properties, a similar duality might be found in the case of matter and the discovery of matter waves by Davisson and Germer and by G.P. Thomson, from which Heisenberg deduced the "uncertainty principle" and Bohr his "principle of complementarity". The 1920s were in fact an heroic age in physics, when many mysteries were clarified.

But many past mysteries remained. And the appealingly simple universe of protons, electrons and photons was almost immediately made more complex by Chadwick's discovery of the neutron, and then by the mesons and by many other particles. And so the search for underlying simplicity has now to be pursued at a deeper level than that of "elementary" particles, with quarks and unified field theories.

The advance of physics during this period depended on the sharpening and augmentation of the senses by instruments. In 1831, even such a simple instrument as the optical lever in the form invented by Poggendorf was only five years old: the commoner form invented by Kelvin did not come until his work on the Atlantic cable in the 1850s.

All the work on gases at low pressure depended on the improving performance of vacuum pumps. There is a direct succession from the experiments on cathode rays right through to the mass

spectrograph and the largest of modern accelerators, which have only been possible because of man's ability to overcome *horror vacui* and make vacuums on a large scale. If there is one invention which more than any other distinguishes science and technology in Western Europe, that is the vacuum pump, which made so much possible.

It is indeed remarkable how much in physics has come out of our attempt to create nothing, just as in mathematics and computation so much has come out of the invention of the sign for zero. And when we realize that all we need in computation beside the sign for zero is a sign for unity, it reminds us that the modern computer is only possible because of an invention made by a physicist, C.E. Wynn-Williams, in 1932 for counting nuclear particles: the scale-of-two counter, which may prove to be one of the most influential of all inventions. The scale-of-two counter depended on the invention of the thermionic valve, which again would have been impossible without vacuum technology.

Old concepts

As late as 1950, Sir Thomas Merton could point out to me that even the latest instruments then in use had not depended on any concept later than that of the electron. More recently, we have seen instruments which do indeed depend on later concepts, the laser being the outstanding example.

Instrumental advances in physics such as the single-isotope lamp, the atomic clock and the laser have led to much increased accuracy in metrology; and in the cases of length and time they have led to standards based on atomic phenomena which could be reproduced even if all material standards were catastrophically destroyed. Such standards are the anchors of a technological civilization.

Physics has had a profound effect on the other sciences. If we regard the object of natural science as understanding the natural universe, the attack is on a broad front. The physicists of the nineteenth and early twentieth centuries made a great breakthrough with quantum mechanics, relativity and atomic theory, resulting in a deeper understanding of both matter and radiation, and in instruments of much greater power and penetration for observing natural phenomena. Through the gap created by the physicists all the other sciences have streamed.

In chemistry, the understanding of atomic and molecular structures and mechanics owes much to the newer concepts of physics. In geology, the interaction of physicists and geologists has led to the development of plate tectonics. Archaeology has advanced through the applications of physical methods of dating. In zoology, the understanding of the capabilities and functioning of the special senses of animals has become even more interesting through the comparison between these senses and those which men have

developed for themselves, for example in maintaining stability through gyroscopes or in means of homing for navigation by IR or by acoustic radar or the polarization of skylight. In physiology, the understanding of the means of transmission of nerve impulses has depended on physical principles; and one of the greatest developments of all, the unravelling of the genetic code, was the joint work of a physicist and a biologist.

Geology versus evolution

Lest we physicists become too conceited, let us remind ourselves of the controversy in the second half of the nineteenth century between Kelvin and the geologists and evolutionists. Kelvin had worked out the ages of the Earth and Sun on the sources of energy generation and loss known in his day, and concluded that neither the Earth nor the Sun could be more than a few hundred million years old⁵. Darwin wanted far longer than that to make his theory of evolution plausible and described Kelvin's work as one of his "sorest troubles".

Kelvin's seemingly patronizing attitude towards the biologists ignored the possibility that there was more in Heaven and Earth than was dreamt of in the natural philosophy of his day. The weakness in his armoury had been pointed out as early as 1869 by T. H. Huxley who said in his presidential address⁶ to the Geological Society:

This seems to be one of the many cases in which the admitted accuracy of mathematical processes is allowed to throw a wholly inadmissible appearance of authority over the results obtained by them. Mathematics may be compared to a mill of exquisite workmanship, which grinds you stuff of any degree of fineness; but, nevertheless, what you get out depends on what you put in; and as the grandest mill in the world will not extract wheat flour from peascods, so pages of formulas will not get a definite result out of loose data.

The weakness was, as Huxley indicated, in Kelvin's assumptions, because he was in no position to take account of nuclear fusion and fission; and the instance is all the more worth remembering because of another of Kelvin's statements (in 1883)⁷ on the importance of converting observations into numbers:

In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be.

Although Kelvin was betrayed by his faith in numbers in estimating the age of the Solar System, this was an almost isolated

exception in a succession of triumphs for exactness in physical measurement. In writing of Joule's work, Kelvin recalled⁸ a meeting at the Royal Society when one Fellow stated that "he did not believe in Joule because he had nothing but hundredths of a degree to prove his case by". But Joule's hundredths of a degree were as crucial as Tycho's 8 minutes of arc, and they triumphantly led to the conservation of energy.

Unfortunately, Kelvin's dictum on measurement had a profound influence on the thoughts of those who accepted it uncritically, and particularly on those working in subjects on the fringes of exact science who wanted to qualify for the scientific fold. If they could measure, they were scientists: if not, they were not. Economics and psychology, and more recently sociology, therefore tried to convert their observations into numbers, sometimes with indifferent success. What they did not realize was that Kelvin had put it the wrong way round, as his attempt at Solar System history showed: it is only when you know a good deal about what you are talking about that you ought to take the step of converting observations into numbers, with the risks that these will become the playthings of politicians and planners. So, in a way, the very success of measurement in physics has led to imitation in subjects unready for it, sometimes with nearly disastrous social consequences.

It is surprising that the importance of physics needs to be stressed at the present time, but we are experiencing a recession of esteem. Physics in defence and in the general economy has received strong support, with the government building up its own establishments at the expense of the universities, industry, and — most unfortunate of all — the schools. Moreover, both government and industry in America were even more appreciative of what physicists might do, and offered substantially more favourable prospects to them, with the result that we lost many physicists in the "brain drain" of the fifties and sixties. One result of all this was that very few good physicists returned to the schools to bring on the new generation; and despite efforts to make the physics syllabus more attractive, the teaching of physics suffered considerably. The government had therefore fostered the paradoxical situation in which, because it realized that physics was so vital, its short-term actions weakened the long-term future of the subject. We have not yet recovered. It is an interesting example of Dean Inge's dictum that "nothing fails like success". □

1. Einstein, A. in *Albert Einstein, Philosopher Scientist* Vol. 1 (ed. Schilpp, P.A.) 53 (Harper, New York, 1959).
2. Rayleigh, Lord *Nature* 43, 26 (1890).
3. Einstein, A. *Nature* 43, 63 (1890).
4. Everitt, C.W.F. *James Clerk Maxwell*, 157 (Scribner's, New York, 1975).
5. Thomson, W. *Proc. R. Soc. Edinb.* 5, 512 (1866).
6. Huxley, T.H. *Q. J. Geol. Soc. Lond.* 25, 38 (1896).
7. Thomson, W. *Lecture to the Institution of Civil Engineers*, 1883.
8. Thomson, W. *Nature* 26, 218 (1882).
9. Bacon, F. *Novum Organum* (1620).

Chemistry: the formative years

George Porter

WITHOUT implying cause and effect, it can be argued that the three decades following the foundation of the British Association in 1831 coincided with the beginning of modern post-Newtonian physics — and of chemistry as a molecular science.

Michael Faraday discovered electromagnetic induction on 29 August 1831, just one month before the first meeting in York. During the same period the particulate description of matter was forced upon reluctant and doubting physicists by the successes of the atomic theory in chemistry: the first president of the committee of chemical sciences of the British Association was John Dalton, whose contributions to atomic theory had been made nearly thirty years earlier. Yet by the time the British Association was founded, chemical theory was in such disarray that few chemists held out any hope that Dalton's "compound atoms" represented anything real or even useful. Here are some typical statements of the time:

Berzelius — Many scientists have cast doubt on the identity of atoms and volumes: but as both theories are only methods of descriptions meant to elucidate the way in which elements combine together, and provided no one pretends that this is what really happens in nature, those theories must be adjudged the best which provide the simplest explanations.

Dumas — What remains of our ambitious incursion into the realm of the atom? Nothing; or at any rate nothing of importance. All we have gained is the conviction that chemistry has gone astray as she always has done when, abandoning experiment, she has tried to walk in the dark without a guide . . . If I had my way, I would expunge the word atom from the vocabulary of science.

How could it be that after the great advances in experimental chemistry by Priestley, Lavoisier and Davy, and after Dalton's apparently clear indication of the way to proceed, chemistry should have reached an impasse?

The reason is, of course, that the physical basis of the atomic theory was non-existent, even contradictory. Even the great Newton had not helped. The development of the atomic theory was hindered by Newton's theory of the pressure of a gas, which he attributed to a repulsive force between the particles proportional to the distance between their centres. Thus, whilst Dalton's model of a compound atom (molecule) was modern, his model of many such atoms in a gas was completely misleading. And Avogadro's hypothesis, proposed correctly in 1811, was taken to refer to atoms rather than to particles (as it must if Newton's view was correct) and this in turn led to an atomic weight of eight for oxygen, six for carbon and to the formula OH for water. It was not until the Karlsruhe conference of 1860 that the terms atom and molecule were finally defined.

Before an atomic theory of matter compatible with commonsense logic as well as with established observations could be developed, two areas of confusion had to be cleared up. The first concerned the distinction between compounds or mixtures, leading to the idea of valency, the structure of molecules, stereochemistry and pointing the way to organic chemical synthesis. The second concerned the nature of heat, which led not only to kinetic theory but to the great laws of thermodynamics.

Molecular weights

The first sentence of the first report of the chemical committee of the British Association in York in 1831 began "It appears to the committee of supreme importance that chemists should be enabled, by the most accurate experiments, to agree in the relative weights of the several elements, hydrogen, oxygen and azote". The committee was to report in the following year, at Oxford. But this committee got no further. Its chairman, James Johnston, reported that "in some cases the evidence in favour of the greater or the less number [of atomic weights] is so equally balanced as to render the decision a matter of great difficulty and one on which the most eminent chemists disagree".

After a discussion on chemical notation, which compared the advantages of the formulae $N + 2C + O$ with the shorter but "algebraically less correct" NC_2O for cyanic acid, the committee regretted that "chemical formulae have not hitherto obtained much favour in England but the state of the science makes their adoption imperative". Gradually chemical formulae were introduced and, rightly or wrongly, they provided for the first time a clear proposition that could be tested.

Inorganic or "mineral" chemistry had by this time been immensely successful in identifying the elements and many classes of their compounds and in establishing the conservation of mass in chemical reactions. The need for a structural theory was not felt. But at the meeting in 1851, the president of the British Association, Charles Dubeny, Professor of Botany in the University of Oxford, said:

A philosopher like Sir Humphry Davy, who had contributed in so eminent a degree to bring the science [of chemistry] into this satisfactory condition, might, at the close of his career, have despaired of adding anything worthy of his name to the domain of chemistry, and have sighed for other worlds to subdue. But there was a world as little known to the chemists of that period as was the Western Hemisphere to the Macedonian conqueror and comprising an infinite variety of important products, called into existence by the mysterious operation of the vital principle, and therefore placed, as was imagined, almost beyond the reach of experimental research.

But now the gulf was being bridged, and it was the new breed of organic chemist who slowly tentatively introduced the ideas of molecular structure and valence. There was no single chemist, no one flash of insight, appearing to illuminate the darkness and confusion. Rather, the truth emerged slowly, the result of the accumulated evidence of many men, and was forced on a generation that seemed obstinately reluctant to formulate the obvious and indispensable theory of valency. These men included Williamson, Gerhardt, Odling, Couper, Frankland, Crum Brown and Kekule, names almost unknown except to other chemists.

In the early 1850s, Edward Frankland was trying to isolate radicals such as methyl and ethyl, groupings which occurred over and over again. He was not successful but he did isolate "a new series of organic bodies containing metals" such as the zinc alkyls. His paper published 10 May 1852, contains the following sentence:

Without offering any hypothesis regarding the cause of this symmetrical grouping of atoms, it is sufficiently evident from the examples just given that such a tendency or law prevails and that, no matter what the character of the uniting atoms may be, the combining power of the attracting element, if I may be allowed the term, is always satisfied by the same number of atoms.

Thus the idea of valency was stated clearly and correctly; Frankland is justly called "the father of the theory of valency". Yet far from having a profound effect on the thinking of the chemists of that time, it caused hardly a ripple of interest. Indeed the secretary of the Royal Society (Professor Stokes) left the paper in his drawer, delaying publication for a year.

Frankland himself was partly to blame for not realizing that the follow through is as important in scientific discovery as in golf. He did not follow up his idea with purely organic compounds, nor did he clearly demonstrate the tetravalency of carbon. This was to come five years later from Kekule and Couper, followed by the first modern-style structural formula from Crum Brown in 1864 and, in the following year, Kekule's ring structure for benzene. So, in 1863, Williamson reported to the British Association "We have now definite and intelligible families of bodies, of which the members are most harmoniously united together by some law of composition and whose connection with neighbouring families is similarly clear and satisfactory".

It was left to Hoffman to make the imaginative leap from formulae on a printed page to molecular models, with croquet balls for atoms and sticks for chemical bonds. He introduced these, and the colour notation still in use, in a discourse before the Prince of Wales on 7 April 1865. But even the most imaginative minds are capable of only one leap at a time. Thus it was that Hoffman copied all his molecular models from the printed page . . . in two dimensions. It was ten years later that Van't Hoff and Le Bel made the

final transition to three dimensions and the tetrahedral carbon atom. Three years before Mendeleev had published his periodic classification of the elements, bringing system to inorganic chemistry and the implication that valency arises from some deep seated structure in the atom itself.

The introduction of the three-dimensional molecular model, with its clearly defined valency rules, heralded the arrival of chemistry as a scientific discipline separate from physics and almost entirely self-sufficient. The developments of these twenty formative years owed little to classical physics or mathematics and henceforth the "natural philosophers" divided themselves more distinctly into one of two camps, chemistry and the rest, which was given (by Tyndall) the new name of physics.

Rutherford's well known pronouncement that chemistry is a branch of physics and the rest of the sciences are stamp collecting (which the Nobel Prize Committee so fittingly rewarded by giving Rutherford the prize for chemistry) missed the point that there is a real difference between the approaches of the chemist and the physicist. Both sciences have been enormously successful. But the single tool most indispensable to the physicist's logic is surely the mathematical equation. The chemist's indispensable tool is the molecular model. When scientists become computers, the physicist computer will be digital and the chemist will be analog. Of course the physicist uses models, but he feels a sense of incompleteness until they have been reduced to mathematical equations. The chemist's experiments usually yield numerical data but, in most of organic and structural chemistry, the objective is to convert the numbers not into mathematical equations but into models.

To avoid further blasphemy I will now declare myself a physical chemist, one who tries to have the best of both worlds. And the years shortly after the foundation of the British Association were the formative years for much of this science also. Chemists have always realized that chemical change is intimately connected not only with change of chemical structure but with the appearance of heat and, often, fire and flame. The closeness of this connection in the chemist's mind is nicely indicated in the little book from which Michael Faraday began his studies, *Conversations in Chemistry* by Mrs Marcet. In

the table of the chemical elements are included heat (caloric) and light. In spite of the statements by Rumford, Young, Davy and others that heat was a mode of motion of the particles of matter, its connection with other forms of energy was obscure.

At British Association meetings between 1840 and 1850, James Joule, a student of John Dalton, described his conversion of mechanical energy into heat and his quantitative measurement of the equivalence between the two. At first they made little impression. But at the 1843 meeting, William Thomson, later Lord Kelvin, was intrigued and excitedly told the others present how important the paper was — although probably incorrect. What Thomson knew, and Joule and the others present did not, was that earlier, a young French engineer, Sadi Carnot, had published a monumental paper on the reverse process, the production of work from heat, with particular reference to the steam engine. In fact, Carnot and Joule were talking about different things . . . Carnot had stated the second law of thermodynamics and Joule the first law.

Thermodynamics emerged at about the same time as structural chemistry, and was again a gradual dawning of understanding, involving many great men among whom Clausius and Kelvin are particularly noteworthy. Its full significance emerged only very slowly and very few understood the entropy concept of Clausius; Kelvin was one who did not. Later were to come the statistical interpretation of thermodynamics and, as late as the 1950s, the connection with information theory. The difficult birth of thermodynamics might be said to be due to the fact that it was a breech delivery; only today, by starting with statistics and information, have we got the thing logically the right way round.

With the relation between heat and energy clear, the way was open to sweep away the remaining difficulties of the Newtonian and Daltonian atoms, for Clausius to introduce the kinetic theory of gases in 1857, and for Loschmidt to calculate the actual number of molecules in a given volume of gas in 1865. So the years between 1831 and 1865 saw the foundation of organic chemistry as a science and the establishment of the two great pillars of physical chemistry, thermodynamics and kinetic theory. As the British Association enters the second half of its second century let us hope that the best years are to come.

The engine we call our Earth

E.R. Oxburgh

THE Greek philosopher Heraclitus maintained as his fundamental axiom *παντα ῥει* — "everything flows". If his words had been heeded by those studying the many diverse aspects of the Earth in the earlier half of this century, a sterile, non-productive and sometimes bitter contro-

versy extending over several decades might have been avoided. The problem was that two groups of scientists with different backgrounds and even different scientific philosophies confronted each other over a fundamental question concerning the behaviour of the Earth: the question

of continental drift.

Many geologists followed the approach of the observational and empirical natural historian. They argued that the distribution of past climates as recorded in the rocks, and the distribution of past faunas documented in the fossil record, were inexplicable if the continents had always been where they are today. It was *just* conceivable that the evidence for large ice sheets in sub-equatorial regions (for example, southern India) could be explained by an extraordinary and inexplicable warping of the Earth's climatic belts, but this could not explain the striking similarities in faunas and evolutionary history observed in land areas now separated by wide oceans. Observations of this kind, and the fact that some of the continents could be fitted together reasonably well as a kind of crustal jig-saw puzzle, led them to accept that during the history of the Earth, continents had moved both with respect to each other and with respect to the poles of rotation. Others, those who approached the Earth from a more physical point of view, reached a very different conclusion. The enormous successes of seismology in the mid-1930s led to the recognition that the patterns of arrival times of seismic waves could be explained almost entirely by an Earth within which seismic velocity increased downwards to about 3,000 km, at which depth there was a sharp decrease. This was the evidence for the Earth's liquid outer core. Equally it was evident that if a major earthquake occurred, the Earth behaved as a gigantic bell struck by a hammer; it simply rang, and the reverberations echoed round the Earth for several hours after it occurred. This was excellent evidence that the outer part of the Earth was strong and rigid.

Seismology

And if the geologists did not believe seismology, what about mountains? Everest was nearly 10 kilometres high and was composed of rock with a density of about 2,850 kg per m³; the rocks at the base of the mountain had to be able to withstand stresses of around 10⁸ N per m², for otherwise they should crack and the mountain would collapse — the height of any structure is limited by the strength of its building materials. This seemed to corroborate the seismological conclusion the Earth was simply too strong for continental drift to occur.

There followed a classic confrontation. With the benefit of hindsight, it is easy to argue that this should never have occurred. The "strength paradox" had been familiar to generations of geologists but had not at that time entered geophysics in an important way. The study of the rocks in some mountain belts had shown the materials observed at the Earth's surface to be hard and strong, and contained clear evidence that in the past they had behaved in a highly ductile fashion or even like viscous fluids.

Another observation might also have caused both geologists and geophysicists to stop and think. The ice of mountain glaciers can behave both as a brittle solid and as a fluid. On time scales of months and years, ice flows slowly downhill in fluid fashion even in some cases displaying the swirls seen in river flow; yet on a short time scale, ice is a hard substance which has to be hammered, sawn or even blasted. If the criteria that had been used to establish the strength of ice had been used to establish the strength of rocks in the continental drift argument, it would have been easy to show that the downhill flow of glaciers under their own weight is impossible.

Clearly something was wrong with the strength arguments. The answer lay in a phenomenon which had been known to engineers and materials scientists for some years — the phenomenon of creep. Creep is observed in materials that are subjected to relatively low stresses for very long times; the materials simply deform continuously, but very slowly, like fluids with an extremely high viscosity. In a crystalline substance, the deformation is accomplished by small-scale processes operating within and between the component crystals and which may involve the diffusion of chemical species through crystals or along grain boundaries, and the change of shape of individual grains by the movement of dislocations within them. All such processes operate most rapidly in materials near their melting point.

It is clear, therefore, that before talking of the "strength" of rocks it is necessary both to specify the time scale and to know something of their temperature. Rocks at the Earth's surface are between 600 and 1,000°C below their melting temperatures and thus creep so slowly that even on geological time scales of millions of years, they may be regarded as solids which are brittle and strong. In the Earth, however, temperature increases relatively rapidly with depth and, below a few tens of kilometres, creep occurs so readily that on time scales of more than a few millions of years, rocks must be considered as fluids even though they are perfectly normal crystalline solids, in the same way as glacier ice.

Therefore the paradox is resolved — the same material may at the same time be "weak" or "strong" — as is the apparent conflict between the seismological and the geological evidence on continental movements. The question remains of how and why these movements occur.

Slow motion

Continents are today moving with respect to each other at rates of a few centimetres per year — about the rate of growth of human fingernails. The surface of the Earth seems to be made up of nearly a dozen internally rigid plates which are in continuous motion with respect to each other. They may slide past each other on major fault zones such as the San Andreas fault of California, or they come together

with one plate overriding its neighbour — this is happening in the Japanese islands — or they may move apart. Where plates move apart, mid-ocean ridges form and hot material wells up from the Earth's interior to occupy the space left as the plates separate. Because the plates are internally rigid, most of the Earth's earthquakes, volcanoes and mountain chains occur along their margins where they interact with each other. The continents are parts of the plates where the crust is unusually thick, and they simply move with the plates to which they belong. The region between the crust (10–70 km thick) and the core is known as the mantle and it is within the uppermost mantle that the effective transition from strong to ductile behaviour occurs. The plates have a capping of crust on top of some tens of kilometres of mechanically coupled mantle.

The only energy source available that is sufficiently large to drive the surface motions is the Earth's heat, which is being continuously lost at a rate of about 60 mW per m² and which is mostly derived from the decay of unstable isotopes of U, Th and K. Rocks are very poor conductors of heat so that if the interior of the Earth is capable of behaving as a very viscous fluid, radiogenic heating ought to be sufficient to set up unstable distributions of density leading to the onset of convection. The effect of convective motions is to allow the Earth to cool more efficiently by moving hot material from the interior towards the surface. The converse of this proposition is telling. If convection did not occur and the only means of heat transfer within the Earth were conduction, the interior of the Earth would probably be largely molten.

So it is tempting to view the Earth as a great vat of molasses or syrup heated in some way from below and slowly churning over with well defined zones of ascending and descending flow, and perhaps a regular and organized arrangement of convection cells within it. Although this is how many familiar fluids behave, it is not the most appropriate model for the Earth's mantle.

For the Earth's heat sources are dispersed throughout the convecting zone rather than beneath it and, moreover, the behaviour of the Earth is dominated by cooling from above rather than heating from below. A better model is to think of the vat of molasses as of nearly uniform high temperature but with a fan blowing cold air across its upper surface. The effect is to chill the near-surface material, making it more dense than the main body of the material below. If at the same time, the material also becomes more viscous, a thin near-surface layer (a so-called boundary layer) is developed that is both stronger and denser than the underlying hot material. This layer tends to sink because it is denser than the material beneath but its higher viscosity and "strength" give it a mechanical coherence which strongly constrains the way it can sink. In fact, the layer would tend to slide as a coherent sheet

into the hot mantle below and new hot material would well up to take its place at the surface, cooling in turn to form a new boundary layer.

This is exactly how the Earth behaves. The surface plates which are continuously in motion are parts of the surface boundary layer, about 100 km thick. Where they converge, one plate slides under another to be resorbed in the Earth's interior; where they separate, as at the mid-ocean ridges, new hot material wells up from below and freezes onto the trailing edges of the diverging plates.

Heat source

The Earth is therefore a great heat engine fuelled by the radioactive decay of unstable isotopes within it. The thermal energy is converted into mechanical energy by the sinking of dense, cold near-surface material and by the rising of hot, more fluid material, to take its place. The surface motions of the plates are simply part of the process by which heat is transferred from the interior to the surface, and the contrasting behaviour of the strong rigid near-surface zones and of the ductile interior reflects the very strong dependence of creep rates on temperature.

This model suggests that the Earth has no permanent crust and that it is all "disposable". Plates move with characteristic velocities of a few centimetres per year and, taking account of the number of plates and the surface area of the Earth, an individual piece of crust should on the average spend about 100 million years at the surface. But one hundred million years is only a few per cent of the age of the Earth and while this age agrees well with what is known of the age of the ocean basins, continental rocks are found in some places to be as old as 3,800 million years, nearly as old as the Earth itself.

New model

So a more elaborate model is needed. Continental crust is three to four times thicker than that in the ocean basins and is composed of low density minerals so that its overall density is less than that of the underlying mantle. The continental parts are therefore never gravitationally unstable, and are not carried back into the interior of the Earth to any great extent. Insofar as most plates have both oceanic and continental parts, the process of subduction (the sliding of a plate back into the mantle) continues until a continental region reaches the point of downturn into the mantle. At that point the buoyancy of the continent prevails and subduction of that plate ceases. When that happens there has to be a worldwide readjustment of plate motions.

It is likely that the motions observed at the surface today have been occurring since the formation of the Earth. In earliest times their character may have been rather different. Rates of movement may have been higher and there may have been more

and thinner plates. But it seems that although the continents have been broken apart or brought together many times by plate motions, they have remained sufficiently buoyant to stay at the surface and not to be carried back into the Earth's interior.

Had this not been the case, suitable environments for the fostering of early life

might well not have developed and we might not have been here today. Even if we had been, we should have found geology a much more speculative subject than it is at present, because our record of the past would have been limited to what could be learned from rocks representing only the most recent few per cent of the history of the Earth. □

The Earth as the home of man

M.J. Wise

IDEAS concerning the influence of environment upon man and of man on environment are deep-rooted. "With the eighteenth century", observes Clarence Glacken "there ends in Western civilization an epoch in the history of man's relationship with nature"¹. Cook's voyages had revealed the essential outline of the map of land and sea, and the time was ripe for the exploration of the interiors of the continents. The Western world was becoming industrialized and spreading its tentacles of trade: technical changes in industry and transport were helping to reshape definitions of resources and the value of locations. New principles of a more specialized science were questioning long-held ideas about the "design" of the Earth. The growth of population itself stimulated thought about man's relationship with nature, not least about the resource capacity of the Earth. Concern with the welfare of people, even if slow to emerge and to translate itself into action, began to show itself.

At the time of the foundation of the British Association the world was still becoming known to Western science. Only unconnected strips of coastline had been traced along the coast of Arctic America. In the Antarctic, Captain Biscoe had discovered Graham and Enderby lands. The *Beagle* was setting out on its momentous journey. Australia was virtually *terra incognita*. Humboldt was back in Berlin, his Russian enterprise which had taken him to the upper reaches of the Ob and the Irtysh, behind him². But in India, the Great Survey was in progress under Sir George Everest's leadership. And in North America the professional surveyors were succeeding the explorers and the frontier was on the move westwards.

It was a world of about one thousand million people. London, the greatest city, had already reached one million. What Lewis Mumford³ has termed the Palaeotechnic phase was gathering strength in Western Europe and North America: coal mines and iron works were changing the countryside, factories were drawing men and women into industrial towns and villages. The railway was in its infancy; the steamship had yet to prove itself.

How, amidst the revolutionary changes that have taken place since then, can we

measure the progress that has been made both in our knowledge of the surface of the Earth and the processes of change, natural and human, at work on it, and in the application of knowledge to the improvement of the conditions of human life? Without doubt we can speak of development and also of evolution. Can we also speak of progress? Until the time of the first World War the idea of indefinite progress had been assumed as an axiom. Today we are less certain.

The problem itself has changed. The world, we say, has shrunk with the advent of air transport, radio communications and orbiting satellites. The exploration of space has caused us to reflect on the limited size of the Earth. Now the people of the Earth are more than four times as many as in 1831 and the prospect of further rapid growth to seven or more million must be faced. Where, we ask, are the resources to be found to feed, clothe, house and employ such an increase?

Meanwhile one fifth of the world's population exists in states of under-nourishment. Great urban megalopolises have risen in North America, Western Europe and the Far East. There is a highly uneven distribution of wealth both within and between countries and the gap between the "developed" and the "developing" world widens annually. The pollution of land, sea and air makes news. In spite of the significant improvements in length of life, states of health and standards of living in many lands, dissatisfaction and community strife fill the headlines in our newspapers.

For the student of relationships between society and the environment, such a period of dynamic change arouses great excitement. It also occasions special difficulties. The problem is no longer only that of observing societies in direct contact with the physical environment. Much of the Earth has now been modified by man and great industrial conurbations provide an artificial environment.

Role of science

The environmental problem has widened. Problems of the environment figure strongly in the programmes of work of most of the subject specializations that have recently emerged. Whether we have yet succeeded in obtaining a satisfactory

integration between disciplines in problem definition, research organization and the analysis of findings is another question.

How can science play its part in assisting the adaptation to new circumstances of societies with long established systems of personal and community relationships, land tenure patterns, forms of political organization? Progress may call for new forms of training which, while providing scientific expertise, also sensitize scientists to the significance of social and economic conditions.

In considering the importance of Darwin's work we no longer deflate the significance of Humboldt, Wallace and others. D. G. MacRae has observed that "Darwin helped to liberate the social sciences and even to create the possibility of their development". Man, it was shown, "might be more than a part of nature but he was certainly part of it". New doors were thus opened for the study of man–environment relationships.

Many attempts during the second half of the century to develop a theory of society–environment relationships exhibited what is now seen as a crude determinism, attempting to establish causal correlations between natural conditions and social behaviour.

At the end of the century, Mackinder's *Britain and the British Seas* was posing the problem of the rise of metropolitan regions in which men and women lived lives detached from nature. "Nowhere", wrote Louis Wirth, "has mankind been farther removed from organic nature than under the conditions of life characteristic of these (great) cities"⁴. So, the influence that city environments exert upon the social life of man became absorbing subjects of study. How do we recreate environments out of decline and decay to meet modern standards of value? Meanwhile, Alfred Weber had begun in 1909 a new strand of thought about the location of industry and the effects of the localization of industry on the growth of industrial conurbations. The discussion of his views and of counter views such as those of August Lösch continues to this day.

Resources

Societies have always been concerned with the problem of resources. The spectre of Malthus has risen again and again. Individual nations have taken steps to conserve stocks or to develop home-based supplies. In our own day the problems of food scarcities in the midst of plenty, the future of energy resources and the supposedly finite nature of Earth resources are raised as major issues. Yet there are many differences between the problem as seen then and now. First is the great extension of our knowledge of the resources. Substances then hardly known are now the bases of major industries. New complex industrial systems based upon the applications of invention, skill and experience provide us with intellectual as

well as material resources. Second the view of the world and its resources has changed. Whereas, in the nineteenth century, economists were concerned with the expansion of the world market, today we have greater concern for global stocks and the rates of use. The idea of "One World" if not yet wholly accepted, finds expression as, for example, in the recent report of the Brandt Commission⁵. Third, resources are, we now understand, not simply substances reposing in or on the Earth but appraisals made by man, arising from his skill, ingenuity and wealth, of what is useful or likely to be useful. And there is, perhaps, if still no cause for complacency, more evidence than there was of mankind's ability to survive.

One of the clearest signs of progress is the widespread acceptance of the idea of conservation. George Perkins Marsh had been writing on "man's alteration of the landscape, intentional and unintentional, desirable and dangerous" since 1847 and his great work *Man and Nature* appeared in 1864⁶. Marsh was not concerned with theory but with the lessons of mis-use derived from experience. By understanding the true impact of his activity man could yet learn "to change the face of the Earth in rational constructive fashion" and so to improve the lot of mankind. Similar ideas were being spread in Europe by A.I. Woeikof and the theme gathered strength.

The influence of themes of conservation as preservation for future use and enjoyment has grown strongly. Political, economic and social have been linked with scientific considerations. However, acceptance of the idea that "conservation is primarily a process of choice" still leaves much room for debate.

The general principles, devised mainly in Europe and North America, have not yet found universal acceptance. Many of the poorer countries, as the United Nations Conference on the Human Environment discovered in 1972, must still give priority to short-term needs. Nevertheless, there are signs of progress. The United Nations Environment Programme was established in 1972. The International Union for the Conservation of Nature and Natural Resources has published its *World Conservation Strategy* (1980) and widespread interest in conservation policies has been demonstrated by governments⁷.

Such strategies display clearly the influence of other ideas, especially the acceptance of the basic principle of ecology. The term was first used in 1868 by Ernst Haeckel but it was not until about 1900 that the study took formal shape. By that time the work of description and classification of climatic conditions as well as of plant life had made possible the publication of Schimper's *Plant Geography* (1898). The development of the concept of ecosystem by Tansley (1935) enhanced the value of ecological approaches in the study of man–environment relationships. The integrating

nature of the concept brought together man and environment in a single framework. It opened the way to studies of structures and of functions. It pioneered the application of systems and model building approaches. The ecological approach has extended far beyond the study of the biological environment.

The work of the sociologist Frederic Le Play (1806–62) the starting point for another influential approach, the study of regionally organized societies. His regional survey method led to the diagnosis of socio-environmental conditions and created a framework which provided a basis for regional planning. The influence can be traced at work in the regional plans of great architect planners like Raymond Unwin and Patrick Abercrombie. Our New Towns and Green Belts present us today with practical outcomes of these ideas. Governments began, haltingly at first, to assume responsibility for the character and quality of urban environments. The pioneers of the approach greatly changed the climate of public opinion. The powerful writings of Lewis Mumford called for the injection of new standards of value into urban and regional designs.

My selection of themes of comment is essentially personal. I cannot exclude from mention the emergence of the concept of the "region". It had many roots, some arising from the need to provide areal classifications of the Earth's surface for the study of plants, climates, population, economies, religions etc. Herbertson's concept of "natural regions" (1905) marked an important step forward. Such scientifically based regions replaced political divisions as units for the teaching of geography. The "region" became an integrating concept for study of relations between society and environment.

More to do

In spite of many signs of progress, much remains to be done. The problems created by population growth, widespread regional inequalities, the existence of deprivation, the varied perceptions that societies hold of their needs, grow in scale year by year. Science itself needs more resources. Its task of creating awareness of the problems and of the vitality of its applications to the improvement of human life is of greater importance than ever before. The problems of integrating approaches from the natural and social sciences and from the humanities have not been fully solved. The study of the human environment has profited from the problem formulation, experimental design and care for measurement of the natural sciences. There has been progress in the adoption of human needs and values in the assessment and application of scientific work.

So the problems of the environment are in part scientific. In part, they are technical, challenging us to improve the standards of living of men without harm to basic resources of land, air and ocean.

They are in part economic, challenging us to count the costs and to form a truer sense of values towards land and landscape. They are in part political, commanding us to combine in improving our institutions and our rules for environmental management. They are in part problems of forecasting, requiring breadth of vision. In matters of the environment, men may in Mackinder's words "be masters of their destiny always provided that they be master of their minds"⁸. And there are many minds for us yet to capture.

Friedrich von Schiller, historian and

1. Glacken, C.J. *Traces on the Rhodian Shore* (Cambridge University Press, 1967).
2. Botting, D. *Humboldt and the Cosmos* (Sphere, London, 1973).
3. Mumford, L. *The Culture of Cities* (Secker and Warburg, London, 1938).
4. Wirth, L. *On Cities and Social Life*, University of Chicago Press (1964).
5. Independent Commission on International Development

poet, observed that "the world's history is the world's judgement". As we survey the world 150 years after the formation of our association, can we be satisfied with the judgment that we make of the Earth as the home of man? We can claim to have made progress in our sciences. Certainly, the problems are more clearly revealed than they were and there is more awareness of them. A distinguished scholar once wrote of the "immeasurable cost of geographical ignorance". In no age has the need for those sciences concerned with the Earth as the home of man been more clear. □

Issues North-South: a programme for survival (Pan, London, 1980).

6. Marsh, George Perkins *Man and Nature*, Scribner, New York (1864).
7. International Union for Conservation of Nature and Natural Resources *World Conservation Strategy*, Gland, Switzerland (1980).
8. Mackinder, H.J. *Britain and the British Seas* (Clarendon, Oxford, 1902).

Exploring the lively Earth 1831–1981

John Sutton

THE more we learn about the Solar System, the more striking becomes the liveliness of the Earth, whose outer parts have been changing for 4,500 million years in response to the flow of heat from the interior. This central concept, which is rooted in Hutton's 200-year-old "Theory of the Earth", opens the way to two lines of enquiry. What changes take place within the Earth and how are they brought about? How can one record the geological history of these changes?

When the British Association first met, these two enquiries had been under way for some 50 years. The discovery that fossils can be used to determine the order in which sedimentary rocks have been deposited, by William Smith in southern England, closely followed Hutton's proof in the late eighteenth century in Scotland that the Earth had a history and had not been created in an instant as biblical tradition had it. But attempts to understand the physical processes at work within the Earth faced intractable problems.

Although Macculloch and contemporaries in the 1820s had realized that an understanding of chemical principles was essential to geology, the physical sciences had not reached the stage where they could explain many geological phenomena. Mathematical modelling of the Earth's interior had, however, begun; Poisson and Fourier had set out to consider temperatures within the Earth and Hopkins (1829) was to show that convection could occur and that solid rock must extend downwards for many hundreds of kilometres. Real progress in this direction had to wait for over a century's advance in physics and chemistry and for identification of slow, but significant geological processes such as sea level

changes or opening and closing of oceans.

The rapid development of a practical method of establishing the history of the fossiliferous sedimentary rocks (but not of the crystalline rocks, whose dating had to wait until the discovery of radioactivity) and the difficulty of putting into practice what Macculloch and others saw was essential, determined the course of nineteenth century geology.

Plate tectonics

Whereas success in arranging the fossiliferous rocks in their correct order contributed to Darwin's understanding of evolution and gave great encouragement to stratigraphical palaeontology, investigations of the physical processes at work within the Earth lost momentum. What followed was a long hard slog: the extension of geological mapping outside Western Europe, the development of new instruments and the patient building up of knowledge of the non-fossiliferous igneous and metamorphic rocks which make up the bulk of the crust. These eventually brought about the general understanding of the behaviour of the crust and mantle, which goes under the name of plate tectonics.

This hypothesis, accepted in the 1960s, stands along with Darwin's work a century earlier and that of Hutton and William Smith at the turn of the eighteenth century as one of the three major advances in geological thought. The timing of these three advances, two achieved in the first 75 years of systematic geological enquiry, the third not until a century later, is illuminating. Such purely geological steps as the recognition of past ice ages, the detection of huge overthrusts within mountain chains and the identification of past faunal provinces, to mention only

three advances, were essential parts of the long study of the Earth which preceded the plate tectonics paradigm.

It is not too difficult to recapture the mood in which geologists approached the first meeting of the British Association. Just as today geological understanding is advancing rapidly in the wake of the grasp of plate tectonics, so in 1831 geology was in turmoil as the thinking of Hutton and Smith was being applied to problem after problem. Murchison and Sedgwick were in the field at the start of their campaign to bring order to the Palaeozoic; incidentally Darwin had his first experience of geological field work during a fortnight with Sedgwick in the summer of 1831. The same year saw the publication of the opening volume of Lyell's *Principles of Geology*, 13,000 copies of which were printed. Darwin was to have a copy in his cabin when he set off in *Beagle* in December of that year. De la Beche had begun the geological mapping of south-west England which led to the setting up of the Geological Survey, the first in the world. The Geological Section of the British Association had a spectacular start and was the largest section at the first meeting, Murchison reported on his field work and, in 1835, jointly with Sedgwick, read the paper establishing the Silurian and Cambrian systems. At the Newcastle meeting in 1938, the membership of Section C had risen to 1,500 and when Sedgwick lectured out of doors on that occasion, 3,000 people attended. The geologists found the new organization particularly suited to their multi-disciplinary needs, while the timing of the meeting in early autumn provided an ideal opportunity to announce the discoveries made that year in the field.

Geologists played a large part in the association as a whole in its early days, when the sciences were interlinked. This wide range of knowledge was precisely what the study of the Earth required to take it forward from the mineral cabinet and the haphazard collections of fossils of the eighteenth century. Lyell gave space to such subjects as vertical movements of the crust and temperature conditions within the Earth in early editions of his text book. De la Beche, in a text *How to Observe Geology* (1836), included instructions for measuring temperatures in the walls of mines and for the manufacture of an instrument to record earthquakes — essentially a tray full of treacle devised by Babbage, who thus anticipated the seismograph as well as the computer.

The very nature of geological enquiry demands organized methods of bringing together the work of individuals. The Geological Society of London, the first of its kind, predated the British Association by a quarter of a century. They provided the forum for debate and for the publication of advances by individual geologists, while the systematic mapping of the British Isles by the Geological Survey

provided the main impetus for advance. By the mid-nineteenth century the Survey extended to Wales and Scotland and so came face to face with novel problems of metamorphic and igneous geology. It was through the Geological Survey that zones of regional metamorphism were first identified in the 1880s, that the field relations of Precambrian gneisses were established in the 1890s and the complexities of British Tertiary volcanoes were set out in a long campaign extending from the late nineteenth century to the 1930s. All these advances originated from geological mapping of a standard which has seldom been equalled.

Framework

During this period the study of crystalline rocks on a worldwide scale was greatly helped by the search for minerals in Precambrian terrain, sparked off by the discovery of gold in South Africa, Canada and Australia, which provided the incentive for systematic mapping of the Precambrian shields. This work provided the framework of knowledge of the older parts of the crust which when radiometric dating became possible in the twentieth century, enabled the geological record to be extended from the base of the Cambrian for 3,000 million years further back in time.

In the mid-nineteenth century, the laboratory study of crystalline rocks took a great step forward when Sorby developed a method of preparing thin sections of rock which could then be examined under the microscope. He announced this method at the British Association meeting in 1958. With this invention it became possible to study the relations of minerals within crystalline rocks and to determine the order in which they had appeared. With the development of experimental petrology, given a great impetus by the founding in 1902 of the Geophysical Laboratory in Washington, this origin of magmas could be investigated. The determination of the internal structure of minerals, made possible by the Braggs' recognition of X-ray diffraction, showed for the first time how the elements are held within the mineral lattices. Since then, Sorby's concern for the detailed examination of individual specimens has been extended to an understanding of geochemical phenomena on a much larger scale. This was a crucial development in visualizing the structure of the Earth.

The attempts to establish temperatures within the Earth in the early nineteenth century and the still earlier study of the Earth's magnetic field mark the start of geophysics. That gravity anomalies at the surface reflected variations of the densities of underlying rocks was appreciated early in the nineteenth century, and led Airy to the notion of isostasy. The British Association played a crucial part with its committee for seismological investigation, which produced the first systematic record

of earthquakes; seismographs were set up as far afield as Cape Town and Western Australia. The first fruit of the systematic study of earthquakes was the delineation early in the twentieth century of the Earth's core, mantle and crust. As the depth of earthquake foci could be established, the remarkable arrangement of earthquakes around the Pacific was recognized by Benioff, and in the 1960s, when the sense of movement in earthquakes could be determined, the full significance of the distribution of earthquakes became clearer: they mark the boundaries of the moving plates.

The study of the magnetism of the Earth, the oldest of the geophysical disciplines, took on new life in the 1950s, when Blackett produced his sensitive spinning magnetometer to measure the weak magnetic fields which many rocks retain from the time they were formed. For the first time it was possible to make geophysical measurement that referred to the geological past. The study of palaeomagnetism produced convincing evidence of the past movements of continents as a result of the work of Runcorn and others in the 1950s. The recognition that the Earth's magnetic field is reversed at irregular intervals opened some remarkable new fields of enquiry. In 1955 Mason and Raff found large magnetic anomalies on the Pacific ocean floor, the origin of which became clear through the work of Matthews and Vine in 1963, who

recognized that the anomalies recorded the cooling of the out-pourings of the volcanic rock which spread from ocean rifts to form the entire ocean floor.

Once it was realized that the crust and upper mantle were in continuous motion, the study of rock deformation took on a new significance. Structural geology or the investigation of deformed rocks began with observations such as those of Daniel Sharpe, on slates and other metamorphic rocks in the mid-nineteenth century; major overthrusts were the first identified at this time. The contact with civil and mining engineering led specifically to the development of the subjects of soil mechanics and rock mechanics. It led also to new methods of investigating how rocks are deformed, which in turn has led to an understanding of rates of deformation within the Earth and thus to our knowledge of plate movement.

Emergence

This broadening of geological enquiry together with the completion for the first time of geological maps of all the continents and the new-found ability both to study the ocean floors and to investigate the deeper crust and mantle brought the Earth sciences to the point where they were equipped possibly for the first time to investigate phenomena on a worldwide scale. It was no accident that plate tectonics, the result of nearly two centuries of investigation, emerged when it did. □

Plate tectonics, biogeography and evolution

A. Hallam

THE link between biogeography and lithospheric plate movements, insofar as they involve the splitting up and collision of continents, is well established but the less direct consequences for sea level and climatic change have been much less thoroughly explored. And although it is a truism that organisms have evolved in response to environmental change, there has been surprisingly little systematic attempt to pursue the possible relationships between major phases of extinction and radiation and the profound environmental changes of plate tectonics.

Plate tectonic processes cause continents in some geographical situations to split along rift zones and to be carried apart passively like giant rafts, while elsewhere they collide with each other. Profound environmental changes can be provoked by such processes. The close relationship between ocean floor topography and distance from spreading axis supports a basaltic cooling model which implies that episodes of rapid spreading increase the volume of buoyant ocean ridge, causing a eustatic rise of sea level. Alternatively, times of extensive continental breakup correlate with an increase in the cumulative

length of active spreading ridges, again causing sea-level rise.

Although not conclusively established, the movement of continents to occupy or surround the poles may have profound global climatic consequences because of the formation of polar ice caps. Comparatively rapid eustatic sea-level oscillations can, of course, result from short-phase climatic fluctuations which affect the ice cap volume. It has been suggested that variations in the albedo with respect to latitude are a result of both the changing distribution of continents and sea-level oscillations, and that the increase in land area in the subtropics in the past 100 million years is the most significant cause of global cooling during that time.

The most obvious effect of continental breakup is to isolate terrestrial organisms by marine barriers, with a consequent increase in endemism. The Triassic reptile faunas are cosmopolitan because free migration was possible across the Pangaea supercontinent, but the Cenozoic mammal faunas of the different continents exhibit significant endemism. What is less obvious is that the distribution of neritic organisms is also severely restricted by wide barriers of

deep ocean, at least if the lifetimes of planktonic larvae are less than the time required to drift across such oceans.

As continents have moved apart, there should be a recognizable *divergence* through time in the similarity of the respective terrestrial and shallow marine organisms. Good examples have been recognized for the late Mesozoic and early Cenozoic faunas of the Americas with respect to Europe and Africa. Conversely, as continents have approached each other and the width of the intervening water barrier has correspondingly shrunk, *convergence* should have taken place. The outstanding example recognized so far concerns the closing of the Iapetus Ocean during the Ordovician and Silurian. Depending on reasonable assumptions about migration potentials of larvae and adults, a plausible story has been worked out of successive migrations across the narrowing ocean of graptoloids, trilobites, brachiopods, ostracodes and finally, after the continental collision, freshwater fish.

A third pattern of change involves the complementary response of terrestrial and marine organisms to continental suturing across a narrow zone. Cross-migration of land organisms gives rise to convergence but the creation of a land barrier causes a formerly homogeneous marine fauna to diverge into two descendent faunas as a consequence of genetic isolation. The emergence of the southern part of the Panama Isthmus during the Pliocene, to link North and South America, is the most celebrated example, and is probably bound up with spreading movements in the Cocos Plate to the west. Somewhat less familiar though probably of even greater biogeographical importance was the early Miocene breakup of the Tethys seaway by emergence of a land barrier in the Middle East. This is bound up partly with impingement of Africa–Arabia on Eurasia and partly with a major sea-level fall in the late Oligocene.

Sea level change

It has not yet been widely appreciated that changes of sea level can also have profound biogeographical effects. Consider, for example, the major late Cretaceous sea-level rise that led to the creation of the North American epicontinental seaway. This extended from the Gulf of Mexico to the Arctic, thereby isolating the dinosaur faunas of the western and eastern landmasses. Late Oligocene withdrawal of the corresponding Turgai Sea of the USSR allowed migration of terrestrial faunas across Eurasia for the first time since the mid Jurassic, whereas North American and west Eurasian faunas had become isolated from each other somewhat earlier by opening of the Norwegian–Greenland Sea sector of the North Atlantic. As regards the neritic invertebrate faunas of epicontinental seas, it is reasonable to expect times of low sea level to correspond with high incidence of endemism because

of physical restrictions to free communication, and supporting evidence can be brought forward from such diverse groups as Cambrian trilobites, Permian fusulines and Jurassic bivalves. Apparently converse results have been obtained, however, for late Cretaceous molluscs, suggesting complicating factors that need detailed examination.

The primary effect of the climatic deterioration that has affected the Earth during the past 50 million years has been to create more temperature-related, latitudinally arranged ecological niches and hence to give rise to an increase in global diversity. The speculation that marine invertebrate faunal provinces in the Palaeozoic were also climatically controlled receives support from the evidence of extensive Ordovician and Permo–Carboniferous glaciation. In at least some examples, however, such as Devonian brachiopods, changes in amount of endemism through time can more plausibly be related to sea-level changes.

Cometary impacts

Whatever the merits of the cometary impact hypothesis to account for the end-Cretaceous extinctions — and many palaeontologists have serious reservations — there is a striking empirical relationship between mass extinction phases in the Phanerozoic and changes of sea level. Supporters of the eustatic hypothesis have generally related the extinctions of neritic faunas to regressions of epicontinental sea provoked by falls of sea level, with a consequent deterioration of the environment as habitats became more restricted, but the problem of why organisms were not able to survive the regressive phases by migrating into refuges has not yet been adequately resolved. An alternative eustatic interpretation sees the succeeding sea-level rise as the more critical factor, because of the frequent association of transgressions with episodes of stagnation (anoxic events), which could rapidly have made benthic and nekto-benthic environments more or less uninhabitable over vast areas. The most likely reason why regressions have adversely affected terrestrial faunas is an increase in continentality of the climate and reduced tolerance to winter cold. Climatic deterioration, involving the formation of the body of cold deep ocean water known as the psychrosphere, is probably the cause of an important late Eocene phase of extinction among foraminifera and ostracodes.

One of the most important evolutionary questions concerns the extent to which mass extinctions affected the faunas randomly or selectively. Was it a case of bad genes or merely bad luck? The available evidence seems to indicate at least some degree of selectivity. Some organisms, such as large vertebrates, ammonites and corals, were relatively susceptible to extinction.

The spectacular adaptive radiations of mammals in the early Tertiary following a long Mesozoic history of stability as a subordinate fauna of small and probably nocturnal animals received its initial stimulus from the relatively sudden vacation of a range of ecological niches by the extinction of the dinosaurs. In other words, biotic factors might initially have been paramount. In contrast, two of the most important radiation phases in the Phanerozoic record may well have been triggered by physical factors ultimately bound up with plate tectonics.

The 'explosive' diversification of Metazoa across the Precambrian–Cambrian boundary corresponds closely to a simple exponential growth model. As the number of taxa increased, the rate of diversification seems to have become diversity dependent. The close correspondence of diversity increase with rise of sea level in the Cambrian suggests the possibility of a causal correlation. Though it cannot be proved, the sea-level rise may well have been a consequence of opening of the Iapetus Ocean, with the growth of a spreading ridge.

The radiation of marine faunas following the massive end-Palaeozoic extinctions was a long-continuing progressive phenomenon, but it is marked by a pulse of acceleration and replacement in the mid Cretaceous, with diversity increase continuing into the Cenozoic. The better known end-Cretaceous extinction event can in many respects be considered as a mere temporary setback in this very dramatic faunal change. The more-or-less coincident, spectacular mid Cretaceous radiations included teleost fish, carnivorous neogastropods and crabs. These changes greatly increased benthic predation pressures and seem to have provoked other changes, such as the decline of surface-dwelling brachiopods and cidaroid sea urchins, the development of protective 'ornamentation' in molluscan prey species and the radiation of infaunal veneroid bivalves and spatangoid sea urchins. There were contemporary radiations of plankton (coccolithophores, foraminifera, diatoms and dinoflagellates) and deep-sea ichnofauna, while on land the angiosperms rapidly replaced the gymnosperms as the dominant plants.

Seafloor spreading

It can hardly be coincidental that these remarkable evolutionary events correspond so closely in time with episodes of exceptional igneous and orogenic activity, the rapid disintegration of Pangaea and the biggest marine transgression since the mid Palaeozoic, apparently produced by a phase of accelerated seafloor spreading in conjunction with a dramatic increase in the length of the oceanic ridge system. Fuller investigation of the interaction of physical and biological events at this time should prove to be a rewarding exercise. □

ARTICLES

A simple model of Saturn's rings

Michel H  non

CNRS, Observatoire de Nice, F-06007 Nice C  dex, France

The distribution of particle sizes in Saturn's rings is probably continuous over many orders of magnitude; there is no 'typical particle size'. A model based on this view accounts for a number of observed properties of the rings: apparent thickness, radar and radio observations, number and size distribution of the gaps discovered by Voyager 1, and optical thickness.

MODELS of Saturn's rings have frequently been based on the convenient assumption that the constituent particles are more or less of the same size (a few authors^{1,2} have considered an extended size distribution). Expressions such as 'the typical particle size', 'the dominant particle size', or even more simply 'the particle size', are commonly encountered in the literature. A slightly less restrictive assumption is that a range of sizes exists, but one can nevertheless define a 'mean particle size' in some meaningful way.

However, natural distributions of particles produced by fragmentation and grinding typically exhibit a continuous distribution of mass over several orders of magnitude³. The notion of a mean particle size is then meaningless. For example, it would not make sense to speak of 'the typical size of an asteroid'.

Saturn's ring particles have certainly been subjected to much fragmentation and grinding, and it thus seems likely that they also exhibit a continuous distribution of sizes, with no preferred value. Here, I describe a very simple model based on this assumption. Computations will be in order of magnitude only; numerical multipliers of order 1 will generally be omitted. Also, the two main rings A and B will be lumped together into a single object, which will be called 'the disk'; the other rings will be ignored.

The size distribution will be defined by the number of particles (for the entire disk) per logarithmic radius increment:

$$F(r) = \frac{dn}{d(\ln r)} = r \frac{dn}{dr} \quad (1)$$

Loosely speaking, $F(r)$ is the number of particles in the range $[r/e^{1/2}, re^{1/2}]$, and can be described as the number of particles with a radius of order r .

Many natural distributions are well represented^{3,4} by a power law

$$F(r) = (r/r_0)^{-\beta} \quad (2)$$

and we shall assume that the disk particles obey such a law. The parameters r_0 and β will be adjusted below to satisfy the observational constraints. β should probably lie within the observed range for particles subjected to grinding³

$$2.4 < \beta < 3.6 \quad (3)$$

Equivalent thickness

A first constraint is provided by the observations of the 'thickness' of the disk; the most recent determinations^{5,6} give 1.4 ± 0.3 km. Actually, what is observed is only the total light received when the disk is edge-on; therefore, the above value should properly be called the 'equivalent thickness', defined as the width of a rectangle having the same area as the disk seen edge-on, and a length equal to the outer diameter of the disk. We assume for simplicity that the centres of all particles lie in the same plane. Figure 1 attempts to show how the edge-on disk really looks; the dotted lines represent the equivalent thickness.

The total projected length of the particles with a radius of order r , in projection on a straight line, is proportional to $r^{1-\beta}$

and increases with decreasing r , if β lies in the range defined by (3). We define a critical radius r' as the value of r for which this total projected length is of the order of the disk diameter

$$2r'(r'/r_0)^{-\beta} = 2R_2 \quad (4)$$

where R_2 is the outer radius of the disk. In the edge-on projection of the disk, particles with a radius of order r fill approximately a rectangle of length $2R_2$ and width $2r$. Smaller particles fall inside the rectangle and do not add to the area. Larger particles protrude out of the rectangle and add some area, but this does not significantly change the total area if $\beta > 2$. Therefore the observed equivalent thickness should be about $2r'$, and we obtain from the observations

$$r \approx 0.7 \text{ km} \quad (5)$$

which, when substituted into equation (4), gives a relation between the model parameters r_0 and β .

Radar and radio observations

Another constraint is provided by the radar and radio observations. We define a second critical radius r'' as the value of r for which the total projected area (of the particles with a radius of order r) is of the same order as the disk area

$$\pi r''^2 (r''/r_0)^{-\beta} = \pi (R_2^2 - R_1^2) \quad (6)$$

where R_1 is the inner radius of the disk. Since the total projected area increases with decreasing r , particles will be collectively 'visible' only if their radius is equal to or smaller than r'' .

On the other hand, observations at a wavelength λ detect only particles with a radius larger than $\lambda/2\pi$ (ref. 7). Thus, positive detection of the disk at a given wavelength implies that $r'' > \lambda/2\pi$; conversely, the absence of detection implies that $r'' < \lambda/2\pi$. (This is an oversimplified description; in practice there are many complicating factors.) A detailed analysis of several observations by Pollack⁷ leads to the conclusion that r'' lies between 2 and 15 cm, and probably closer to the lower value. A similar analysis by Cook and Franklin⁸ gives $r'' \approx 8$ cm. We shall adopt here

$$r'' = 4 \text{ cm} \quad (7)$$

The two parameters r_0 and β of the size distribution can now be computed from relations (4)–(7). Taking $R_1 = 90,000$ km and $R_2 = 137,000$ km, we obtain

$$r_0 = 32 \text{ km}, \quad \beta = 3.2 \quad (8)$$

The value of β falls well within the range (3). Thus, with a continuous distribution of sizes there is no problem in satisfying simultaneously the two observational constraints (5) and (7). We remark that neither r' nor r'' can in any sense be considered as a typical size, or even as some sort of average size; as shown by equations (4) and (6), they result from combinations of the parameters (r_0, β) of the size distribution with the dimensions of the disk. Essentially, r' and r'' are the radii at which particles are in sufficient numbers to cover, respectively, a one-dimensional or a two-dimensional projection of the disk.

Jeffrey's limit

The size of the largest particles should be below Jeffrey's limit⁹, which is the maximal size for which a solid particle can resist tidal disruption. From equations (1) and (2), the number of particles with a radius larger than r is found to be

$$n(r) = \beta^{-1}(r/r_0)^{-\beta} \quad (9)$$

The approximate radius of the largest particle can be taken to correspond to $n = 1/2$; for $\beta = 3.2$, this radius is then $0.86r_0$, or 28 km. This agrees with the classical value of 100 km for Jeffrey's limit.

Mass of the disk

For $\beta > 3$, the total mass of the disk is

$$M_D = \int_{r_{\min}}^{\infty} \frac{4}{3} \pi r^3 \rho \, dn = \frac{4\pi}{3(\beta-3)} \rho r_0^{\beta} r_{\min}^{3-\beta} \quad (10)$$

where ρ is the density of the particles, assumed to be spherical. It is necessary here to assume a lower cutoff $r_{\min}$ to avoid divergence. $r_{\min}$ is not known; fortunately, it enters equation (10) with a small exponent. If we take $\rho = 1 \text{ g cm}^{-3}$ and $r_{\min} = 1 \text{ cm}$, we obtain $M_D = 1.4 \times 10^{22} \text{ g}$; if $r_{\min} = 0.1 \text{ cm}$, $M_D = 2.2 \times 10^{22} \text{ g}$. It might be significant that these values for the mass of the disk are of the same order as the masses of the first two satellites of Saturn—Mimas ($4 \times 10^{22} \text{ g}$) and Enceladus ($8 \times 10^{22} \text{ g}$).

Radial structure

We turn now to the dynamics of the disk. We assume circular and coplanar particle orbits. As shown by Goldreich and Tremaine^{10,11}, particle orbits in a disk system repel each other; at each close approach resulting from conjunction, some angular momentum is transferred from the inner particle to the outer particle. The rate of angular momentum transfer can be computed by a perturbative approach when the orbits are sufficiently distant¹¹; in order of magnitude, it is

$$\dot{\Phi}_{12} \sim G^2 m_1 m_2 (m_1 + m_2) \omega^{-2} d^{-4} \quad (11)$$

where G is the gravitation constant, m_1 and m_2 are the masses of the two particles, $d = a_2 - a_1$, a_1 and a_2 are the radii of the orbits, and ω is an average angular velocity

$$\omega^2 = GM_S a^{-3} \quad (12)$$

M_S is the mass of Saturn and $a = (a_1 + a_2)/2$ is an average orbital radius.

The angular deflections of the two particles, in the system of their common centre of mass, are of the order of

$$G(m_1 + m_2) V^{-2} d^{-1} \sim G(m_1 + m_2) \omega^{-2} d^{-3} \quad (13)$$

where V is the relative velocity of the two particles. Relation (11) breaks down when these deflections are no longer small, that is, for $d \leq d_0$, with

$$d_0 = [G(m_1 + m_2) \omega^{-2}]^{1/3} = [(m_1 + m_2) M_S^{-1}]^{1/3} a \quad (14)$$

Assuming that the density of the disk particles is of the same order as the density of Saturn, we have

$$m_i r_i^{-3} \geq M_S a^{-3} \quad (15)$$

where $i = 1, 2$, and r_1 and r_2 are the radii of the two particles. On the other hand, the fact that the ring lies inside the Roche limit corresponds to

$$m_i r_i^{-3} \leq M_S a^{-3} \quad (16)$$

Combining these relations, we obtain

$$d_0 \sim (r_1^3 + r_2^3)^{1/3} \sim \max(r_1, r_2) \quad (17)$$

For $d \leq d_0$, the interaction of the two particles is described by Hill's case of the restricted three-body problem. Numerical integrations show a complex interplay, sometimes with a temporary capture of the two particles. The rate of angular momentum transfer varies erratically with d_0 , but remains of the same general order of magnitude for $0 \leq d \leq d_0$. This order of magnitude is obtained by substituting $d = d_0$ in equation (11)

$$\dot{\Phi}_{12} \sim G m_1 m_2 d_0^{-1} \quad (18)$$

Collisions between the particles tend also to increase their radial separation¹². They operate only for $d < r_1 + r_2$, that is, essentially for $d \leq d_0$. They produce a rate of angular momentum transfer which can be shown to be of the order of relation (18). Adding the effects of gravitational interaction and of collisions, we find that the total rate of angular momentum transfer is still given by equation (11) for $d \geq d_0$, and by relation (18) for $d \leq d_0$.

We hypothesize now that because of this repulsive effect, a sufficiently massive particle will be able to push away the other particles, both towards the inside and towards the outside, and thus to clear completely an annular zone. This would explain the numerous dark gaps observed by Voyager 1¹³. We assume that this happens above a critical particle radius r_c . The particle population is then divided into: (1) particles with $r > r_c$, which we shall call heavy particles; each of them describes an isolated orbit inside the gap which it has created; and (2) a 'continuous' distribution of particles with $r < r_c$, or light particles; actually, this distribution is separated into several ringlets by the heavy particles. We compute now the critical radius r_c .

We assume for simplicity that there is no mass segregation inside the continuous distribution. Then the number of light particles with radius r and orbital radius a is

$$F(r) g(a) d(\ln r) da \quad (19)$$

where $F(r)$ is given by equation (2) for $r < r_c$ and is zero for $r > r_c$, and $g(a)$ describes the radial distribution of particles; such that

$$\int g(a) da = 1 \quad (20)$$

We compute now the flux of angular momentum Φ_a across a circle of radius a . We consider first inner particles with a given radius r_1 and outer particles with a given radius r_2 . As shown above, two such particles interact essentially when their radial separation is of the order of d_0 ; and the rate of angular momentum transfer for one pair is of the order of relation (18). Therefore, the rate of angular momentum transfer for all pairs is of the order of

$$F(r_1) g(a) d(\ln r_1) d_0 F(r_2) g(a) d(\ln r_2) d_0 G m_1 m_2 d_0^{-1} \quad (21)$$

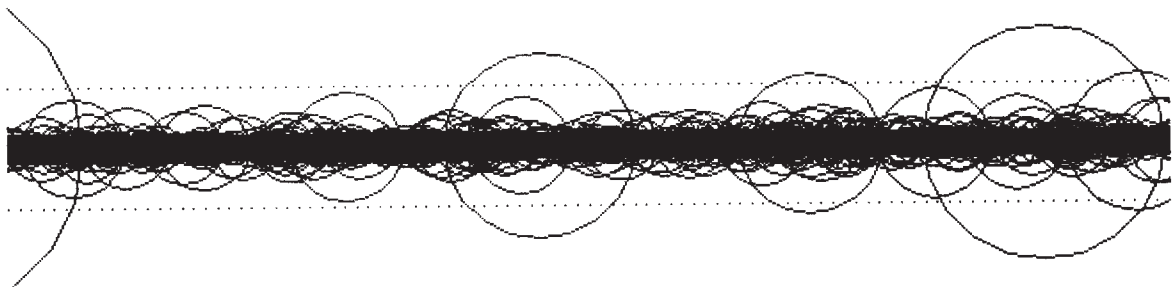


Fig. 1 A close-up view of the disk seen edge-on, obtained by numerical simulation. The centres of the particles have been randomly placed on the horizontal axis with a uniform distribution; their radii have been randomly chosen in accordance with the distribution law (2), with $\beta = 3.2$ and $r_0 = 32 \text{ km}$. The length of the strip shown is 12 km, that is, an extremely small part of the whole disk. Dotted lines represent the equivalent thickness (see text). Particles are represented as if they were transparent.

Next, we integrate over all radii r_1 and r_2 , on the interval $r_{\min} < r_1 < r_c$. The result is

$$\Phi_a \sim G\rho g^2(a)r_0^\beta M_D r_c^{4-\beta} \quad (22)$$

Consider now a thin annulus of radial width da . If $g(a)$ is constant, this annulus loses to the outside the same flux (22) which it receives from the inside. If, on the other hand, there is a gradient in $g(a)$, then the annulus gains angular momentum at a rate of the order of

$$-G\rho g \frac{dg}{da} da r_0^\beta M_D r_c^{4-\beta} \quad (23)$$

We suppose now that there exists a heavy particle of radius r_1 on a circular orbit of radius a_1 , and we ask how the presence of this particle modifies the continuous distribution in its neighbourhood. For definiteness, we shall consider the outside distribution, that is, $a > a_1$; symmetrical considerations apply to the inside distribution. For $d > d_0$, a light particle in the outside region receives angular momentum from the heavy particle at a rate given by equation (11). An annulus of radial width da and average radius a therefore receives angular momentum at a rate of the order of

$$G^2 m_1 g(a) da \omega^{-2} (a - a_1)^{-4} \int_{r_{\min}}^{r_c} m_2 (m_1 + m_2) F(r_2) d(\ln r_2) \quad (24)$$

Since $r_1 > r_c > r_2$, this gives approximately

$$G^2 m_1^2 g(a) da M_D \omega^{-2} (a - a_1)^{-4} \quad (25)$$

We assume that the disk has reached a state of quasi-equilibrium, in which its distribution does not change with time (except on a very long time scale, of the order of the age of the disk). Then the total angular momentum received by an annulus, that is, the sum of relations (23) and (25), should vanish. This gives

$$\frac{dg}{da} \sim r_0^{-\beta} r_c^{\beta-4} r_1^6 (a - a_1)^{-4} \quad (26)$$

which can be integrated:

$$g \sim g_0 - r_0^{-\beta} r_c^{\beta-4} r_1^6 (a - a_1)^{-3} \quad (27)$$

g_0 is the value of the density g at a large distance from the heavy particle, that is, the unperturbed density. If Q is the fraction of the disk occupied by ringlets and $L = R_2 - R_1$ is the radial width of the disk, then from relation (20) we have

$$g_0 \sim (QL)^{-1} \quad (28)$$

The second term in equation (27) represents the perturbation by the heavy particle. It is negative the presence of the heavy particle produces a local depletion of the disk density. At a distance $a - a_1 = d = d_0$, we have from equations (17), (27) and (28)

$$g \sim (QL)^{-1} - r_0^{-\beta} r_c^{\beta-4} r_1^3 \quad (29)$$

It can be verified that for $d < d_0$, g remains essentially of the order of equation (29). Thus, this equation represents the minimal value which can be reached by g . The heavy particle will clear a gap if this minimal value is negative. Therefore, the critical value r_c is the value of r_1 for which equation (29) equals zero

$$r_c \sim [r_0^\beta (QL)^{-1}]^{1/(\beta-1)} \quad (30)$$

Using the values (8), $Q = 0.5$ (see below) and $L = 47,000$ km, we obtain

$$r_c \sim 1.6 \text{ km} \quad (31)$$

The number of heavy particles is

$$n(r_c) = \beta^{-1} (r_c/r_0)^{-\beta} \sim 4,500 \quad (32)$$

This agrees reasonably well with the number of gaps observed by Voyager 1¹³.

The width w of the gap created by a heavy particle of radius r is twice the value of $a - a_1$ for which relation (27) vanishes; using relation (30), one finds

$$w \sim r^2 r_c^{-1} \quad (33)$$

Thus, the width of a gap is proportional to the square of the radius of the particle. This result has also been obtained by Goldreich¹⁴. For the largest particle, with a radius of the order of 30 km, we obtain $w \sim 600$ km, again in order-of-magnitude agreement with Voyager 1 observations¹³. At the other end, particles with a radius of the order of r_c produce a gap whose width is itself of the order of r_c . Voyager 1 did, in fact, observe gaps down to its limiting resolution of 2 km (ref. 15).

Optical thickness

The total width of all gaps is $(1 - Q)L$; it can also be found by summing relation (33) over all heavy particles

$$(1 - Q)L \sim \int_{r_c}^{\infty} r^2 r_c^{-1} (r/r_0)^{-\beta} d(\ln r) \sim r_0^\beta r_c^{1-\beta} \quad (34)$$

The main contribution to this total comes from the many narrow gaps with a width close to the lower limit, that is a few kilometres.

Substituting relation (30) in relation (34), one finds

$$(1 - Q)L \sim QL \quad (35)$$

Thus, Q should lie somewhere inside the $[0, 1]$ interval without being close to either 0 or 1. This result is independent of all physical parameters: the disk settles automatically into a state where the gaps and the ringlets occupy comparable areas.

The observed optical thickness τ of the disk is of order unity throughout most of the A and B rings. The traditional interpretation of this fact has been that the total projected area of the particles must be of the same order as the disk area, that is, that $r_{\min} \sim r''$ —a remarkable and unexplained coincidence. Here we arrive at a completely different interpretation, by assuming only that $r_{\min}$ is smaller than r'' . The ringlets are then practically opaque, and light can pass only through the gaps. τ is then given by

$$e = 1 - Q \quad (36)$$

and is automatically of order unity.

Conclusion

This model is admittedly very crude. Only orders of magnitude have been considered. Many important effects have been omitted—resonances with satellites, ellipticity and non-planarity of particle orbits, Poynting-Robertson and other non-gravitational effects. A very simple mass distribution has been used. Finally, it is an aggregate model, which lumps together Saturn's rings in spite of their obvious differences. Nevertheless, it is perhaps significant that even a very simple model seems to be able to account for the main properties of the rings, once the crucial assumption of a continuous distribution of masses has been made.

I thank A. Brahic, P. Goldreich and H. Rickman for useful comments on a first draft of this paper.

Received 23 April; accepted 28 July 1981.

- Greenberg, R., Davis, D. R., Hartmann, W. K. & Chapman, C. R. *Icarus* **30**, 769 (1977).
- Cuzzi, J. N., Durisen, R. H., Burns, J. A. & Hamill, P. *Icarus* **38**, 54 (1979).
- Hartmann, W. K. *Icarus* **10**, 201–213 (1969).
- Dohnanyi, J. S. *J. geophys. Res.* **74**, 2531–2554 (1969).
- Brahic, A. & Sicardy, B. *Nature* **289**, 447–450 (1981).
- Fountain, J. W. & Larson, S. M. *Icarus* **36**, 92–106 (1978).
- Pollack, J. B. *Space Sci. Rev.* **18**, 3–93 (1975).
- Cook, A. F. & Franklin, F. A. in *Planetary Satellites* (ed. Burns, J. A.) 412–418 (University of Arizona Press, Tucson, 1977).
- Jeffreys, H. *Mon. Not. R. astr. soc.* **107**, 260–262 (1947).
- Goldreich, P. & Tremaine, S. *Nature* **277**, 97–99 (1979).
- Goldreich, P. & Tremaine, S. *Astrophys. J.* **241**, 425–441 (1980).
- Brahic, A. *Astr. Astrophys.* **54**, 895–907 (1977).
- Smith, B. A. *et al. Science* **212**, 163–191 (1981).
- Goldreich, P. Contr. 3545 (Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, 1980).
- Kelly Beatty, J. *Sky Telescope* **61**, 7–18 (1981).

A gene product required for correct initiation of segmental determination in *Drosophila*

Gary Struhl

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The properties of mutations of the homoeotic gene extra sex combs (esc) indicate that the product of this gene is required early in development for correctly initiating segmental determination. Specifically, the esc^+ gene product appears to be a necessary component of the process by which other homoeotic genes, such as those of the bithorax complex, are selectively turned on, or off, in particular segmental primordia. The resulting combinations of active and inactive genes established at this time then maintain the specific pathways of development followed subsequently by the different segments.

THE insect epidermis is composed of a series of unique head, thoracic and abdominal segments. Each segment is formed by the descendants of a group of cells set aside at the blastoderm stage¹⁻⁵, and hence is a developmental compartment^{6,7}. In *Drosophila*, several homoeotic genes have been identified which appear to control the specific developmental pathways followed by the different segments⁸⁻¹⁰. Mutations in these genes generally lead to the transformation of entire segments into other segments¹⁰.

Most homoeotic genes examined previously seem to be required throughout development to maintain particular determined states (reviewed in refs 11, 12). Here I describe a maternal-effect, homoeotic gene whose product is essential early in development for initiating the correct determined state in each segmental primordium. Specifically, the product of this gene, *extra sex combs*, may be required for selectively turning on, or off, the genes of the *bithorax* complex and possibly other homoeotic genes as well.

Mutations of the *extra sex combs* gene

The *extra sex combs* (*esc*) locus¹³⁻¹⁵ was originally defined by a single recessive mutation (esc^1) mapping on the second chromosome. esc^1 homozygotes develop into adult flies in which the distal portions of the second and third legs are partially transformed into first legs. This transformation is most apparent in the second and third legs of males which develop sex comb bristles normally found on only the first leg. The *extra sex combs* phenotype caused by the esc^1 mutation is variable, both in penetrance and expressivity^{13,15}, suggesting that this mutation might reduce, but not eliminate, the gene product. Consequently, to assess the role of the esc^+ gene it was first necessary to obtain mutations which would completely inactivate it.

New mutations of the *esc* locus were obtained as follows. Appropriately marked wild-type males were mutagenized with ethylmethane sulphonate or γ rays and outcrossed to a stock heterozygous for the original esc^1 mutation. Male progeny bearing the mutagenized chromosome *in trans* to the esc^1 chromosome and showing an *extra sex combs* phenotype were selected and outcrossed to isolate the mutagenized chromosome. Eight new mutations ($esc^{3,4,5,6,7,8,9,10}$) were obtained, including a small deficiency (esc^{10} : deficient for the salivary chromosome bands 33B1,2). In addition, a new spontaneous mutation (esc^2) was recovered. Several of these new alleles ($esc^{2,4,5,6,8,9,10}$) have the same, extreme phenotype whether homozygous or *in trans* to each other, suggesting that they eliminate most, if not all, the wild-type activity of the locus. In the following experiments, these apparent null alleles were used interchangeably and are referred to simply as *esc*.

Maternal effects of *esc* mutations

When $esc^-/+$ mothers are fertilized by *esc* bearing sperm, the esc^-/esc^- zygotes can develop into adult flies, even though the

wild-type gene itself is absent following fertilization. In contrast, when esc^-/esc^- zygotes are obtained from esc^-/esc^- mothers, a dramatic homoeotic transformation is observed. Instead of developing into adults, these zygotes develop into first instar larvae in which all the abdominal and thoracic segments, and some of the head segments, are transformed into eighth abdominal segments (Fig. 1). Not surprisingly, these larvae fail to hatch from the egg case and die. Because the presence of the wild-type gene during oogenesis alone is sufficient to rescue this homoeotic phenotype, it follows that the esc^+ gene product derived from the mother persists (or perdures¹⁶) long enough after fertilization to allow the correct development of almost all the segments.

The homoeotic phenotype of embryos derived from esc^- mothers depends critically on the genotype of the zygote. If eggs from esc^- mothers are fertilized by sperm bearing one copy of the esc^+ gene, they show a range of intermediate phenotypes in between the normal larva and esc^- larvae derived from esc^- mothers (Fig. 2). Briefly, the thoracic and abdominal segments appear partially, or completely, transformed into eighth abdominal segments whereas the head segments appear only partially transformed or normal. If eggs from an esc^- mother are fertilized by sperm bearing a duplication of the esc^+ gene, they frequently develop into adults (Fig. 3). However, even though these paternally rescued adults have carried two copies of the wild-type gene throughout development, more than half show a patchy transformation of one or more abdominal or thoracic segments into more posterior segments. For example, patches of haltere tissue are sometimes found in the wing. Similarly, in males, patches of darkly pigmented cuticle characteristic of the fifth and sixth abdominal segments are sometimes found in the first four segments of the abdomen (Fig. 3). Thus, absence of the esc^+ gene product in a newly fertilized egg can be partially corrected by introduction of the wild-type gene via the sperm. Moreover, the extent to which normal development is restored depends on the number of copies of the wild-type gene introduced, two copies being sufficient in some cases to result in normal development.

Role of the *esc* gene product in activating genes of the *bithorax* complex

The phenotypes resulting from mutations of the *esc* locus suggest that the product of the wild-type gene is required for initiating the correct determined state in most, if not all, of the segmental primordia. A simple hypothesis for the role of the esc^+ gene is that its product is normally required early in development for selectively activating (or inactivating) other homoeotic genes. These genes remain active (or inactive) during subsequent development and thereby specify the particular developmental pathways followed by the different segments.

The complex locus, *bithorax*, provides a good candidate for a set of homoeotic genes whose members require the esc^+ gene product for their correct activation. On the basis of mutant

phenotypes, Lewis¹⁰ has proposed the following model of the *bithorax* complex. The complex contains a series of at least eight genes which normally promote particular thoracic and abdominal pathways of development as opposed to a ground state, mesothoracic pathway^{8,10}. The arrangement of these genes on the chromosome corresponds directly with the arrangement of segments in which they are active, and in which they specify the correct developmental pathway. Thus, in the mesothorax, none of the genes is active, in the metathorax the first few genes are active, in the first abdominal segment the next gene(s) is also active, and so on, until the eighth abdominal segment, in which all the *bithorax* genes are active.

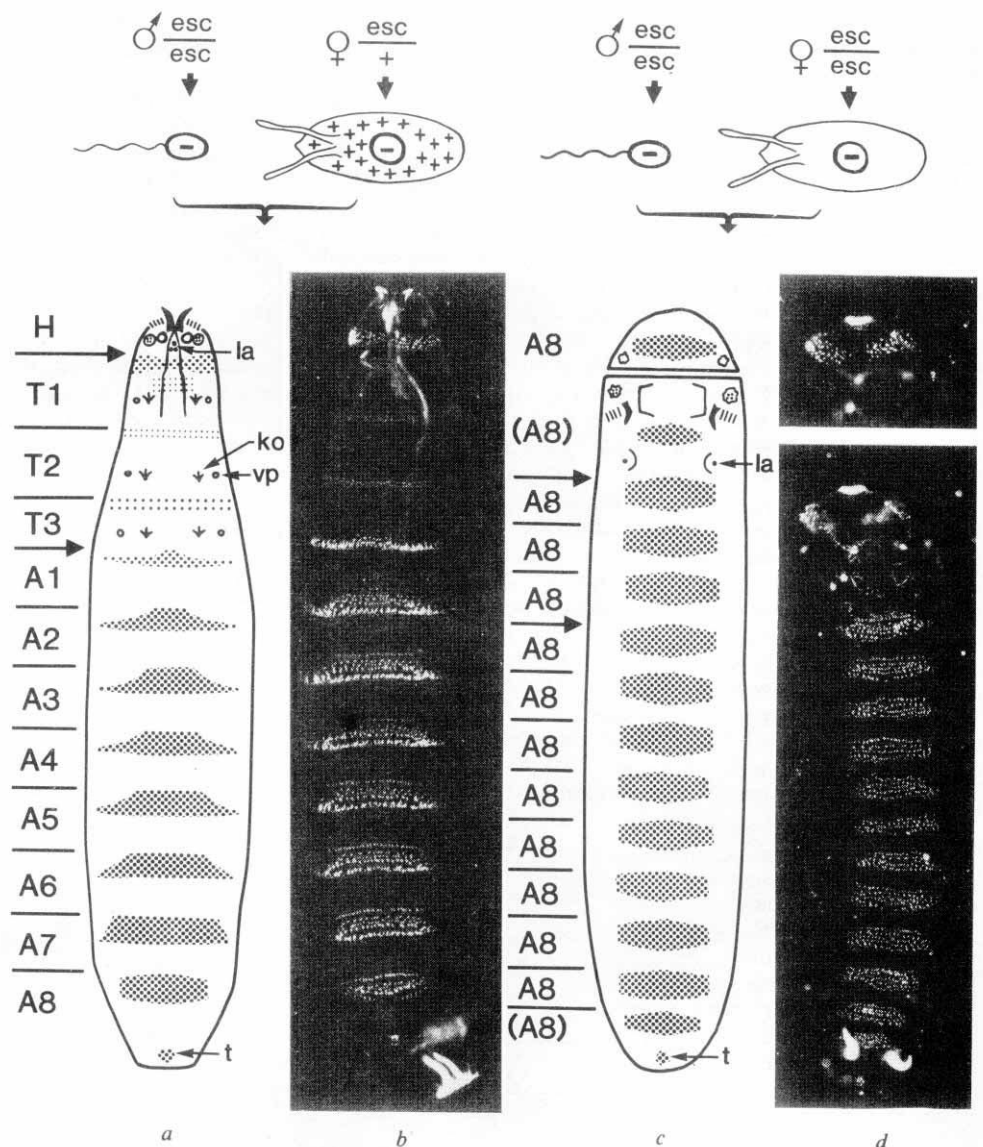
Thus the phenotype caused by mutations of the *esc* locus, in which nearly all the segments appear transformed into eighth abdominal segments, may result from an indiscriminate activation of all the *bithorax* genes. To test this hypothesis, the phenotypes of *esc*⁻ and *esc*⁺ embryos in which different portions of the *bithorax* complex had been deleted were determined.

Four aberrations of the *bithorax* complex (*Tp(3)bx^d100*, *T(2;3)P10*, *DfUbx¹⁰⁹* and *DfP9*; see ref. 10) have been used alone or in combination to generate a series of deletions which

remove particular portions of either the centromere-proximal or centromere-distal side of the complex (Fig. 4). The phenotypes of these deletions are described in detail in Fig. 5, and agree largely with previously published descriptions¹⁰. Deficiencies which delete proximal portions of the complex transform most, or all, of the abdominal segments towards a mesothoracic segment in a graded fashion. For example, *DfP10* (Fig. 4) results in the complete transformation of the metathorax and first abdominal segment into mesothoracic segments, and in a partial transformation of the remaining abdominal segments towards mesothoracic segments such that each segment appears more thoracic and less abdominal than the segment posterior to it (Fig. 5). In contrast, deficiencies which delete distal portions of the complex appear to transform segments discretely. For example, when only the proximal portion of the complex carried by *DpP10* (Fig. 4) is present, the metathorax and first abdominal segment appear normal but the remaining seven abdominal segments are transformed into first abdominal segments (Fig. 5). None of these deletions affects the development of the mesothorax, prothorax or head segments.

The phenotypes of these deletions in embryos lacking the *esc*⁺

Fig. 1 Maternal effect of mutations of the *esc* locus. The phenotype of *esc*⁻/*esc* larvae arising from *esc*⁻/*esc*⁺ mothers (a, b) and *esc*⁻/*esc*⁻ mothers (c, d) are shown. The body of the normal larva consists of three thoracic and eight abdominal segments each having a characteristic pattern of hairs and sensory organs, and of several head segments which form specific sensory organs and cuticular structures clustered around the mouth (see refs 22, 23 for detailed descriptions of the hair patterns and sensory organs of the first instar larva). The larvae shown here have been fixed, cleared and mounted as in ref. 24. This procedure allows all of the cuticular features of the larva to be discerned using phase contrast, Nomarski or dark field optics. It is therefore possible to identify unambiguously the characteristic patterns of all the thoracic and many of the abdominal segments in wild-type and homeotically transformed larvae. a, b, *esc*⁻/*esc*⁻ zygotes arising from *esc*⁻/*esc*⁺ mothers develop into normal larvae which give rise to virtually normal adults. Both the diagram and the photograph show primarily ventral cuticular structures of the larva. Only the characteristic patterns of ventral hairs are apparent in the photograph—some of the sensory organs unique to the thoracic segments (the Keilin's organ, ko, and ventral pit, vp) are indicated in the diagram. A tuft of hairs, t, is found posterior to the eighth abdominal segment; its segmental status is unknown^{22,23}. The paired posterior spiracles appear posterior to the tuft in the photograph. c, d, *esc*⁻/*esc*⁻ zygotes arising from *esc*⁻/*esc*⁻ mothers develop into larvae in which all the thoracic and abdominal segments as well as one or more of the head segments bear the rectangular band of ventral hairs and unique sensory organs^{22,23} (not apparent in the photograph) characteristic of the eighth abdominal segment. The transformed segments frequently lack or only partially form posterior spiracles (another characteristic of the eighth abdominal segment), indicating that the transformation is often not complete. c, d Show primarily the ventral side of the larva except for the uppermost portions of each which show the dorsal side of the head. Note that the head fails to undergo the normal process of involution so that remnants of the sensory organs and other cuticular derivatives lie on the outside of the larva, anterior to the transformed thoracic segments. One segment of the head is consistently transformed; curiously, this segment forms a band of hairs characteristic of the ventral, eighth abdominal segment on the dorsal side of the head. Note also that 12 transformed segments are found posterior to the remnants of the labial segment, whereas in wild-type embryos, only 11 segments (3 thoracic and 8 abdominal) are apparent. This additional segment probably corresponds to a cryptic ninth abdominal segment in the normal larva. Abbreviations: H, head segments; T1, T2 and T3, pro-, meso- and metathorax; A1, A2 ... A8, abdominal segments one to eight; la, labial sense organ; ko, Keilin's organ; t, tuft; vp, ventral pit. The boundaries between the head and thorax and between the thorax and abdomen are marked by arrows. × 100.



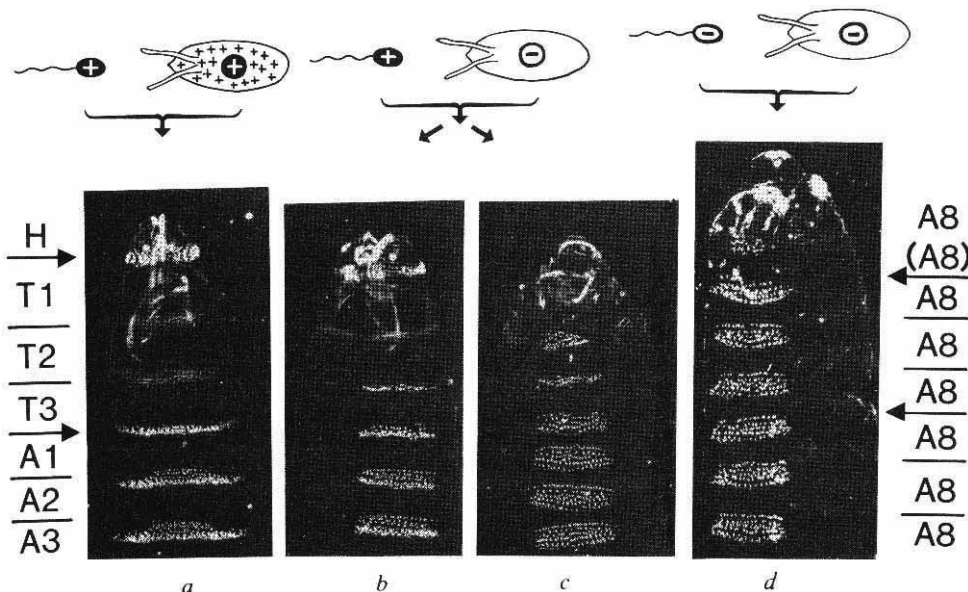


Fig. 2 Partial rescue of the *esc* phenotype by sperm carrying one copy of the *esc*⁺ gene. *esc*⁻/*esc*⁺ zygotes arising from *esc*⁻/*esc*⁻ mothers develop into larvae in which most, or all, of the thoracic and abdominal segments appear partially transformed into more posterior segments. Anterior halves of wild-type (a), partially rescued (b, c) and *esc*⁻ (d) embryos are shown. Note that the thoracic and abdominal segments of b and c are transformed to different extents (for example the metathorax of b carries a thin band of abdominal hairs whereas the metathorax of c bears a rectangular band of thick hairs similar to that of the normal eighth abdominal segment). Segments are indicated as in Fig. 1. $\times 90$.

Discussion

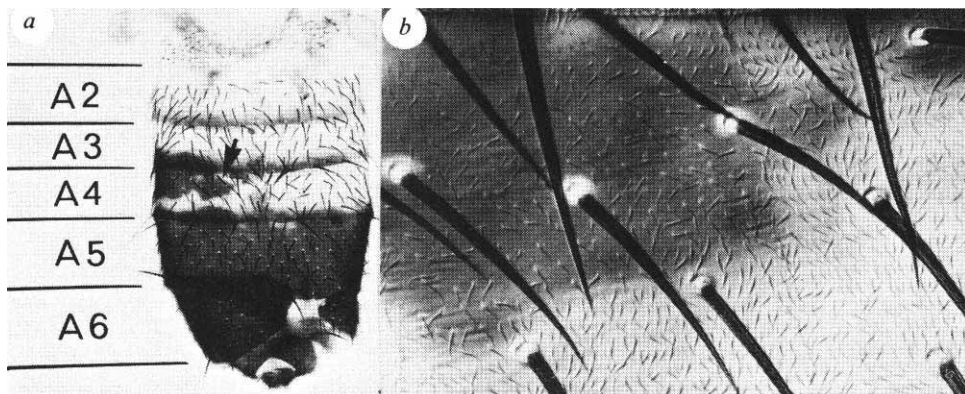
Mutations of the *esc* locus, unlike all previously described homoeotic mutations, result in a dramatic homoeotic transformation only when both the mother and the zygote are mutant. They can therefore be described as maternal-effect, homoeotic mutations. Previously described maternal-effect mutations which alter the segmental pattern of the embryo do so by altering the number or polarity of segments (for example, *bicaudal*¹⁷, *fused*¹⁸). The existence of such mutations reinforces the conclusions of other studies¹⁹ that the mother provides the egg with a system of spatial information which is necessary for establishing the characteristic number, sequence and polarity of the segments. In contrast, mutations of the *esc* locus alter only the characteristic patterns of the different segments, not their number or polarity. Thus, in addition to supplying V the egg with spatial information, the mother appears to provide gene products which are essential for the correct interpretation of this information during embryonic development.

Two lines of evidence indicate that the *esc* gene product has an early role in the determination of segments. First, when both the mother and the zygote lack the *esc*⁺ gene, most, or all, of the segments develop as eighth abdominal segments. However, this homoeotic phenotype can be completely rescued if the mother carries a copy of the wild-type gene, indicating that the presence of maternally derived gene product is sufficient to allow normal segmental determination in mutant embryos. Unless the maternally derived gene product persists, and hence rescues, the mutant phenotype throughout development, this result suggests

gene product (Fig. 5, bottom row) appears, to a first approximation, to obey a simple rule. All the thoracic and abdominal segments, as well as a segment of the head, appear to develop like the eighth abdominal segment formed in *esc*⁺ embryos carrying that same *bithorax* deficiency. Thus, the state of the *bithorax* complex (that is, whether wild type or partially or completely deleted) seems to determine the developmental pathway followed by the eighth abdominal segment. In the absence of the *esc*⁺ gene product, all the thoracic and abdominal segments follow this same pathway of development.

The phenotype of *esc*⁻ embryos homozygous for *Dfp9*, which is thought to delete the entire *bithorax* complex¹⁰, is worthy of special note. If the *esc*⁺ gene product is normally required for the correct activation of only the *bithorax* complex, the development of embryos lacking the entire complex should be the same in the presence or absence of the *esc*⁺ gene product. However, this is not the case (Fig. 5). In *esc*⁺ embryos homozygous for *Dfp9*, the head, pro- and mesothorax develop normally but the metathorax and first seven abdominal segments are transformed into mesothoracic segments and the eighth abdominal segment is transformed into an intermediate between the pro- and mesothorax. In contrast, all the thoracic and abdominal segments of *esc*⁻; *Dfp9* embryos seem to develop as intermediates between the normal pro- and mesothorax (that is, similar to the eighth abdominal segment of *esc*⁺; *Dfp9* embryos); moreover, at least one head segment develops in part like a thoracic segment. Assuming that *Dfp9* does indeed remove the entire *bithorax* complex, this result suggests that the *esc*⁺ gene product is also required for correct activation of other homoeotic loci.

Fig. 3 Partial rescue of the *esc* phenotype by sperm carrying two copies of the *esc*⁺ gene. *esc*⁻/*esc*⁺ zygotes arising from *esc*⁻/*esc*⁻ mothers frequently develop into adults which show a patchy transformation of one or more thoracic and abdominal segments into more posterior segments. a Shows the abdomen of a paternally rescued, male fly. The fourth segment carries a patch of dark cuticle (arrow) characteristic of the cuticle of the normal fifth segment. Note that the cuticle of the transformed patch (b) resembles that of the fifth segment not only in colour but also in being creased and having fewer hairs than the cuticle of the fourth segment. Sperm carrying two copies of the *esc* gene were obtained using Dp(2;3)17 (Ising, personal communication) in which the region 31B1-34D1.2 (Ashburner, personal communication) is translocated to the left arm of the third chromosome. Segments are indicated as in Fig. 1. a, $\times 40$; b, $\times 340$.



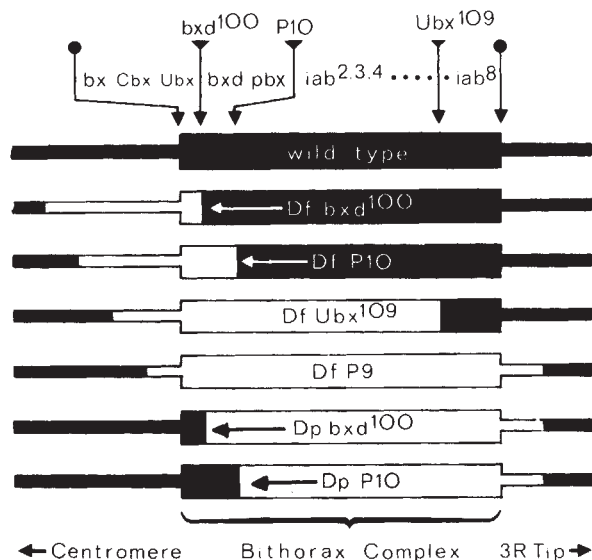
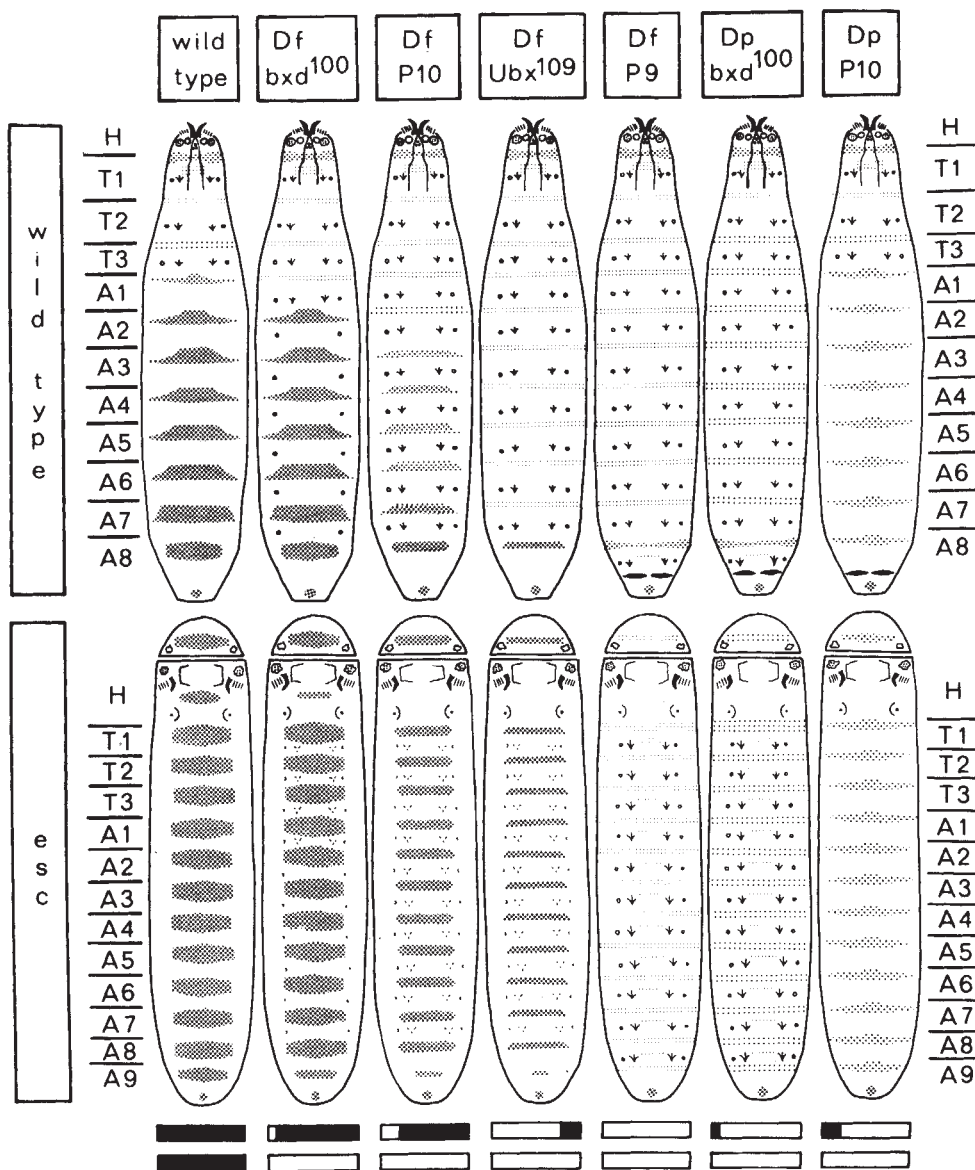


Fig. 4 Deletions of the *bithorax* complex. Four aberrations (*Tp(3)bxd*¹⁰⁰, *T(2;3)P10*, *DfUbx*¹⁰⁹ and *DfP9*; described in detail in ref. 10) have been used alone and in combinations to construct a series of deletions of the *bithorax* complex. *DfP9* is thought to delete the entire complex. *DfUbx*¹⁰⁹ deletes most of the centromere-proximal portion of the complex leaving only a small, centromere-distal fragment. *Tp(3)bxd*¹⁰⁰ transposes a small, proximal portion of the complex to the left arm of the third chromosome. This fragment, referred to as *Dpbxd*¹⁰⁰, can be separated from its reciprocal deficiency, designated *Dfbxd*¹⁰⁰. *Dpbxd*¹⁰⁰, in the absence of any other copies of the *bithorax* complex, is equivalent to a deletion of the distal portion of the complex. *T(2;3)P10* translocates a somewhat larger, proximal portion of the complex to the left arm of the second chromosome. Like *Tp(3)bxd*¹⁰⁰, the proximal and distal fragments created by *T(2;3)P10* can be separated to give *DfP10*, a simple deletion of the proximal portion of the complex and *DpP10*, which in the absence of other copies of the *bithorax* complex is equivalent to a deletion of the complementary, distal portion. The relative positions of the breakpoints of *Tp(3)bxd*¹⁰⁰, *T(2;3)P10* and *DfUbx*¹⁰⁹ have been inferred from their complementation behaviour with other mutations of the complex; a map of the genes identified by these mutations as well as of the breakpoints of the four aberrations is shown at the top of the figure (see refs 8, 10). The portions of the complex which are absent in each deletion are shown in white; the portions of the locus which are present are shown in black.

Fig. 5 The phenotypes of deletions of part, or all, of the *bithorax* complex in *esc*⁺ and *esc*⁻ embryos. The role of the *esc*⁺ gene product in activating the genes of the *bithorax* complex has been tested by removing part, or all, of the complex in both *esc*⁺ and *esc*⁻ embryos. The figure is a matrix in which the top row shows the phenotypes of *esc*⁺ embryos carrying all, part or none of the *bithorax* complex, and the bottom row shows the phenotypes of *esc*⁻ embryos carrying the same *bithorax* genotypes. The *bithorax* genotype of each column of embryos is stated at the top of the figure and shown systematically as in Fig. 4 at the bottom (for example, the *Dfbxd*¹⁰⁰ embryo carries *Dfbxd*¹⁰⁰ in trans to *DfP9*). Similarly, the *esc* genotype of each row of embryos is stated on the left. The wild-type embryo in the upper left corner shows the characteristic pattern of each segment (see Fig. 1) and serves as a key for illustrating the homeotic phenotypes of the remaining embryos. Note that in each *esc*⁻ embryo, most of the segments appear identical. Note also that the segment reiterated in a particular *esc*⁻ embryo appears similar, or identical to the eighth abdominal segment of the *esc*⁺ embryo positioned above it, and having the same *bithorax* genotype. (The anterior portions of the reiterated segments of *esc*⁻; *DfP9* and *esc*⁻; *Dpbxd*¹⁰⁰ embryos are different from the eighth abdominal segments of *esc*⁺; *DfP9* and *esc*⁺; *Dpbxd*¹⁰⁰ embryos, respectively. However, in both cases, the pattern of the reiterated segment as a whole is intermediate between the normal prothorax and the meso- or metathorax, just as in the eighth abdominal segment of their respective *esc*⁺ embryos.) In many cases the pattern of a particular transformed segment seems to be intermediate between two or more segments of the normal larva. In *esc*⁺ embryos which are *DfP9*, *Dpbxd*¹⁰⁰ and *DpP10*, two patches of sclerotized cuticle appear posterior to the eighth abdominal segment (see also ref. 10); these are shown as dark bands. In *esc*⁻ embryos which carry *Dfbxd*¹⁰⁰, *DfP10* and *DfUbx*¹⁰⁹, the Keilin's organ and ventral pits (see Fig. 1) are frequently absent or poorly formed.



that the *esc* gene product is critically required only early in development. Second, the absence of the *esc*⁺ gene product in eggs from mutant mothers can be partially rescued by wild-type sperm, and when the sperm carries two copies of the wild-type gene, the resulting zygotes can develop into adult flies. These adults appear normal except for occasional patches of transformed tissues in the thorax and abdomen. One interpretation of this patchy phenotype is that a critical amount of the *esc*⁺ gene product is required at the time segments are determined; if this amount is not reached in all cells, then some may be incorrectly determined and hence give rise to transformed tissue in the adult. That this apparent failure of cellular determination is not corrected during subsequent development, even though two copies of the wild-type gene are present throughout, suggests that the *esc*⁺ gene product has a critical role specifically at the time when segments are determined.

The extra sex combs phenotype of *esc*⁻ adults is not consistent with the *esc*⁺ gene product having an exclusively early function in development. Homozygous *esc* flies derived from heterozygous mothers show a transformation of the distal portions of the second and third legs into distal portions of the first leg. This same phenotype has been observed in clones of *esc*¹/*esc*¹ cells induced in *esc*¹/+ larvae¹⁵. Thus, the *esc*⁺ gene seems to function later in development in some of the imaginal disks of the larva. The relationship of this late function to the early role of the gene product is unclear.

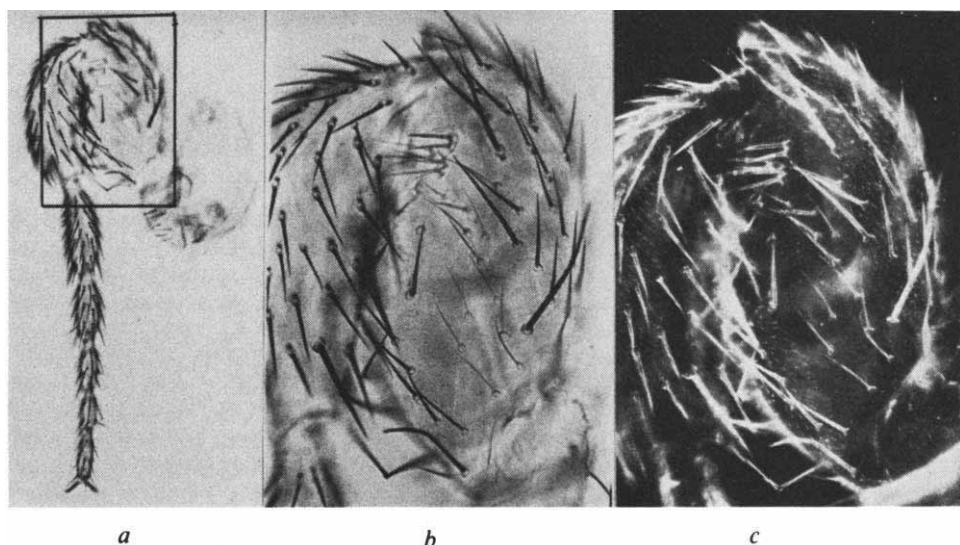
The results of deleting part or all of the *bithorax* complex in *esc* and *esc* embryos provide several insights into the role of the *esc* gene product. In general, absence of the *esc*⁺ gene product results in an embryo in which most of the segments appear identical, suggesting that this product is necessary for initiating the unique developmental pathway followed by each of the different segments. The particular segmental type reiterated in *esc* embryos depends critically on the state of the *bithorax* complex. In embryos carrying all of the *bithorax* genes, eighth abdominal segments are formed; in embryos lacking a particular part of the *bithorax* complex, the reiterated segment appears similar, or identical, to the eighth abdominal segment of *esc*⁺ embryos lacking that same part of the complex. The simplest interpretation of this result is that all the extant *bithorax* genes are normally active in only the eighth abdominal segment, but that in the absence of the *esc*⁺ gene product, all these genes become active in most or all of the segments. Thus, the *esc*⁺ gene product appears to be an essential component of the process which specifies the correct combinations of active and inactive *bithorax* genes in each of the segmental primordia. Finally, *esc*⁻

embryos lacking the entire *bithorax* complex reiterate a thoracic-like segment throughout the embryo. Because this homoeotic phenotype differs significantly from that of *esc*⁺ embryos lacking the entire complex, it suggests that the *esc*⁺ gene product is also required for the correct activation of homoeotic loci outside the *bithorax* complex.

The phenotype of mutations of the *esc* locus resembles that observed in embryos homozygous for recessive lethal mutations of the *Polycomb* (*Pc*) locus¹⁰, on the third chromosome. On the basis of this embryonic phenotype, as well as the phenotype of *Pc* embryos lacking the entire *bithorax* complex, Lewis¹⁰ has proposed that the *Pc*⁺ gene product normally acts as a repressor of the *bithorax* genes. The similarity between the homoeotic phenotypes of *esc*⁻ and *Pc* embryos suggests that mutations of both loci disrupt the process of segmental determination by altering the expression of the *bithorax* genes. The evidence presented here indicates that the *esc*⁺ gene product may be essential only early in development for the correct activation (or inactivation) of the *bithorax* genes. In contrast, two lines of evidence together indicate that the *Pc*⁺ gene is required throughout development. First, homozygous *Pc* embryos obtained from heterozygous mothers show a homoeotic transformation, indicating that the *Pc*⁺ gene is normally required during embryonic development¹⁰. Second, clones of homozygous *Pc* cells induced throughout the larval period in the thoracic and eye-antennal disks appear to differentiate an alia-like tissue (Fig. 6), whereas clones in the analia appear relatively normal. This result indicates that the *Pc*⁺ gene is required throughout the larval period for the proper development of the thoracic and head segments. Thus, it is possible that the *esc*⁺ gene function is required for initiating, and the *Pc*⁺ gene function for maintaining, the correct determined state in each segment.

Determination has been defined by Hadorn²⁰ as "a process which initiates a specific pathway of development by singling it out from among the various possibilities for which a cellular system is competent", and as a process which results "in a reproducible cell state propagated in the nondifferentiating phase by cell heredity". These elements of determination have been described in genetic terms in the selector gene hypothesis proposed by Garcia-Bellido^{12,21}. In this model, the continuous activity of particular combinations of selector genes maintains the developmental pathways followed by all the cells giving rise to specific compartments. The process by which these developmental pathways are initiated is thus tantamount to the process by which the correct combinations of active and inactive

Fig. 6 Homozygous *Pc*² clone induced during larval development. *a, b, c* Show a female, first leg bearing a *Pc*²/*Pc*² clone at $\times 60$, $\times 175$ and $\times 175$ magnification, respectively (dark field optics are used in *c*). The clone is marked with the bristle colour mutation yellow, *y* (apparent in *b*). Note that the marked bristles lack bracts (small, triangular projections at the base of most leg bristles) and tend to cluster together. In addition, the cuticle formed by the clone is devoid of hairs and is therefore apparent in *c* as a swath of bald cuticle running through the background of wild-type leg tissue which carries an even covering of hairs. All these features are characteristic of cuticle found on the analia and parts of the male genitalia. Similar *Pc*² clones are found in the thoracic segments and eye-antenna of both males and females. Also, *Pc*² clones in the genitalia of both males and females disrupt the normal structure, whereas *Pc*² clones in the analia are relatively normal. Thus, *Pc*² clones seem to transform the thoracic segments as well as the eye-antenna into analia-like tissue. The particular clone shown here was induced by X-irradiation during the late first or early second instar of a *y; ju trc Pc*²/*Dp*(1; 3)*sc*¹⁴, *y*⁺ *M*(3)*i*⁵⁵ larva. (See ref. 25 for descriptions of the *y*, *ju trc* and *M*(3)*i*⁵⁵ mutations and of the *y*⁺ duplication, *Dp*(1; 3)*sc*¹⁴, *tricorned* (*trc*) is a recessive mutation which causes single hairs to develop as tufts of small hairs (A. Garcia-Bellido, personal communication); it maps ~0.1% to the left of the *Pc* locus which is positioned on the left arm of chromosome III, close to the centromere.) Three classes of clones were observed following X-irradiation of larvae of this genotype: (1) *y* clones which were *trc*⁺, *Pc*⁺ and developed normally, (2) rare, *y*, *trc* clones which developed normally, and (3) *y* clones which formed analia-like tissue devoid of hairs. The distribution of clones in these three classes indicates that mitotic recombinations occurring between the *Pc* locus and the tip of the left arm of the *ju trc Pc*² chromosome produce *y* clones which develop normally, whereas recombinations between the *Pc* locus and centromere give rise to clones which form analia-like tissue.



selector genes are established in the founding cells of each compartment.

The process of segmental determination in insects can be interpreted according to the selector gene hypothesis. Because segments are precise regions of the epidermis formed exclusively by the descendants of specific cell groups¹⁻⁵, they are, by definition^{6,7}, compartments. The properties of the *bithorax* genes⁸⁻¹⁰ are consistent with their being selector genes which normally maintain the particular developmental pathways followed by the metathorax and first eight abdominal segments. The product of the *esc*⁺ gene appears essential early in development for selectively turning on, or off, the *bithorax* genes, and possibly, other selector genes. Thus, it may provide

the first example of a gene product which is required for initiating the correct combinations of active and inactive selector genes in the segmental primordia.

I thank P. A. Lawrence for his advice and encouragement, D. A. Melton for interesting discussions, C. Nüsslein-Volhard and E. Wieschaus for teaching me about embryos, A. Garcia-Bellido, G. Ising, G. Jürgens, E. B. Lewis, G. Morata and T. R. F. Wright for sending me essential strains, G. Richards for ascertaining the cytology of *esc*¹⁰, and several of my colleagues in the Cell Biology Division of the Laboratory of Molecular Biology for critical reading of the manuscript. I also thank the Nuffield Foundation and the MRC for financial support.

Received 6 March; accepted 2 June 1981.

1. Lawrence, P. A. *J. Embryol. exp. Morph.* **30**, 681-699 (1973).
2. Wieschaus, E. & Gehring, W. *Dev. Biol.* **50**, 249-269 (1976).
3. Steiner, E. *Wilhelm Roux' Arch. Entw. Mech. Org.* **180**, 9-30 (1976).
4. Lawrence, P. A. & Morata, G. *Dev. Biol.* **56**, 40-51 (1977).
5. Szabad, J., Schüpbach, T. & Wieschaus, E. *Dev. Biol.* **73**, 256-271 (1979).
6. Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature new Biol.* **254**, 251-253 (1973); *Dev. Biol.* **48**, 132-147 (1976).
7. Crick, F. H. C. & Lawrence, P. A. *Science* **189**, 340-347 (1975).
8. Lewis, E. B. *Am. Zool.* **3**, 33-56 (1963); in *The Role of Chromosomes in Development* (ed. Locke, M.) 231-252 (Academic, New York, 1964).
9. Morata, G. & Garcia-Bellido, A. *Wilhelm Roux' Archiv. Entw. Mech. Org.* **179**, 125-143 (1976).
10. Lewis, E. B. *Nature* **276**, 565-570 (1978).
11. Morata, G. & Lawrence, P. A. *Nature* **265**, 211-216 (1977).

12. Garcia-Bellido, A. *Am. Zool.* **17**, 613-629 (1977).
13. Slifer, E. H. *J. exp. Zool.* **30**, 31-40 (1942).
14. Hannah-Alava, A. & Strommaes, O. *Drosoph. Inf. Serv.* **29**, 121-123 (1955).
15. Tokunaga, C. & Stern, C. *Dev. Biol.* **11**, 50-81 (1965).
16. Garcia-Bellido, A. & Merriam, J. R. *Proc. nat. Acad. Sci. U.S.A.* **68**, 2222-2226 (1971).
17. Nüsslein-Volhard, C. in *Determinants of Spatial Organisation* (eds Subtelney, S. & Konigsberg, I.R.) 185-211 (Academic, New York, 1979).
18. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
19. Sander, K. *Wilhelm Roux' Arch. Entw. Mech. Org.* **151**, 430-497 (1959); **151**, 660-707 (1960); *Adv. Insect Physiol.* **12**, 125-238 (1976).
20. Hadorn, E. *Brookhaven Symp. Biol.* **18**, 148-161 (1965).
21. Garcia-Bellido, A. in *Cell Patterning*, 161-182 (Ciba Foundation, North Holland, Amsterdam, 1975).
22. Lohs-Schardin, M., Cremer, C. & Nüsslein-Volhard, C. *Dev. Biol.* **73**, 239-255 (1979).
23. Turner, F. R. & Mahowald, A. P. *Dev. Biol.* **68**, 96-109 (1979).
24. Van der Meer, J. *Drosoph. Inf. Serv.* **52**, 160 (1977).
25. Lindsley, D. L. & Grell, E. L. *Carnegie Inst. Wash.* **627**, (1968).

LETTERS TO NATURE

Grand unification, the neutron electric dipole moment and galaxy formation

John Ellis*, Mary K. Gaillard†, D. V. Nanopoulos* & Serge Rudaz*

* Cern, CH-1211 Geneva 23, Switzerland
† Lapp, 74019 Annecy-le-Vieux Cédex, France

Grand Unified Theories (GUTs) provide a possible solution to the long-standing cosmological problem of the apparent asymmetry between matter and antimatter¹⁻⁹. We have argued recently¹⁰ that in a class of GUTs one may relate the baryon asymmetry to certain contributions to the neutron electric dipole moment, d_n . In the absence of cancellations due to an unforeseen symmetry, we thereby obtain a qualitative cosmological lower limit on d_n . We emphasize here that the experimental upper limit on d_n and the present cosmological baryon-to-photon ratio (n_B/n_γ) can be used to limit the amount of extra entropy generated after baryosynthesis. This restriction on entropy generation becomes quite severe if, as we argue here on the basis of explicit GUTs, d_n is actually considerably larger than our previous cosmological lower bound. As an application of our restriction of entropy generation, we establish an upper limit on shear in the early Universe.

The mechanism for baryon generation proposed by GUTs leads^{11,12} to adiabatic perturbations ($\delta T/T \approx \frac{1}{3}(\delta\rho/\rho)$) in a Robertson-Walker-Friedmann (RWF) universe. However, astrophysical observations tend to favour isothermal perturbations as the origin of galaxies and other large-scale structures in the Universe (for reviews, see refs 13, 14). One constraint is the apparent quadrupole anisotropy recently reported^{15,16} in the microwave background radiation, which in an adiabatic model would probably imply¹⁷ dipole anisotropy in excess of observation¹⁸, though this discrepancy might be reduced¹⁷ if there were massive neutrinos. Although adiabatic perturbations are not excluded as the origin of galaxies, it becomes very important to understand whether isothermal perturbations are compatible with the GUT mechanism for baryon generation. Several scenarios for creating isothermal perturbations have been proposed^{11,12,19,20} by deviating from a strict RWF cosmology. These include the explosions of primordial black holes (PBHs) which

themselves originate from adiabatic perturbations^{11,12,19} and inhomogeneities in curvature, vorticity, the cosmological constant or shear^{19,20}. Alternatively, isothermal fluctuations could arise in an RWF model when CP is violated¹⁹ softly at a GUT phase transition preceded by an inflationary de Sitter epoch²¹.

We object to other scenarios and focus on inhomogeneities in shear as the most plausible sources of isothermal perturbations. Any isothermal modes generated by PBH explosions would probably be overwhelmed by the growing adiabatic modes associated with the formation and explosion of the holes. Curvature fluctuations would have to be uselessly small²⁰ to avoid the domination of the present Hubble expansion by scalar curvature, and they would in any case give rise to growing adiabatic perturbations which would overwhelm the isothermal perturbations¹⁹. Fluctuations in the cosmological constant could occur during a phase transition in the early Universe, but they would presumably be restricted to the size of the horizon at that time, and hence much smaller than the scale of a galaxy²⁰. Such fluctuations could be extended if the causally connected domains were blown up by a de Sitter expansion during a first-order GUT phase transition, but it is difficult to assess this possibility in the absence²¹ of a convincing mechanism for terminating the transition. The soft CP violation proposal¹⁹ relies on a similar idea of expanding exponentially a small sub-horizon region to become a large region which then breaks CP spontaneously in a correlated way. Apart from the problem of boundaries between domains where CP is violated differently, we object to this proposal because in a model of spontaneous CP violation one cannot know during the symmetric phase how CP will be violated subsequently. Some, as yet unknown, mechanism has to be invented for 'encoding' the instructions for the mode of breakdown before the de Sitter phase starts, storing the information during the inflationary period and then 'decoding' the message in parallel in a myriad of separated domains. The same hurdle has to be crossed if one is to suppress monopole production in an inflationary universe²¹. For an alternative argument for suppression of magnetic monopole production, see ref. 22.

In view of these considerations, we shall concentrate on inhomogeneous shear^{19,20} as a possible source of isothermal fluctuations. This may be the most natural and attractive possibility, as there is no direct evidence that the RWF metric is

applicable in the very early Universe. One is naturally led to consider more general anisotropic models like the simple Bianchi type I cosmologies in which there is primordial shear. As the baryon asymmetry generated depends on the local expansion rate, which in turn depends on the ratio $\Sigma \equiv (\sigma^2/8\pi G_N \rho)$ of shear to the canonical energy density²⁰, fluctuations in these quantities are related by:

$$\delta\left(\frac{n_B}{n_\gamma}\right)/\left(\frac{n_B}{n_\gamma}\right) \sim \delta\Sigma/\Sigma \quad (1)$$

for large shear Σ , and $\sim \delta\Sigma$ for small shear Σ . One might question whether shear large enough to generate interesting isothermal fluctuations could exist during baryon number generation when the temperature $T \sim 10^{15}$ GeV. In addition to the normal diminution of shear, $\Sigma(T) \sim T^2$, dissipative processes due to the grand unified viscosity²³ associated with the long mean free paths of elementary particles at temperatures $\geq 10^{15}$ GeV might wash out some of the shear before the baryon number was generated, or might dilute the baryon-to-entropy ratio by creating more entropy after the baryon asymmetry was established. Preliminary estimates²³ of the grand unified viscosity have been roughly confirmed²⁰, indicating that all the shear is not dissipated before baryon generation, but that an extra entropy factor

$$E^{4/3} \approx [10^{-5} \Sigma(m_X)^{-1/2}] \quad (2)$$

can be created after baryon generation²⁰. The present non-zero $(n_B/n_\gamma) \geq 2 \times 10^{-10}$ (ref. 24) means that the entropy factor E , and hence the shear $\Sigma(m_X)$, cannot be too large²⁰. As the original CP-violating baryon asymmetry must be $\leq 0(\alpha/\pi)$, one can deduce from equation (2) that $\Sigma(m_X) \leq 0(10^{29})$. This limit on shear is already many orders of magnitude better than the present best limit from nucleosynthesis^{25,26} which leads to $\Sigma(m_X) \leq 0(10^{36}-10^{42})$. However, it is possible to improve this limit, using our results on the neutron electric dipole moment.

We have recently¹⁰ argued that in a class of GUTs with no unanticipated cancellations, diagrams analogous to those giving (n_B/n_γ) also contribute to renormalization of the quantum chromodynamic (QCD) θ parameter and hence indirectly to the neutron electric dipole moment d_n . They give an order of magnitude lower bound

$$d_n \geq 0(2.5 \times 10^{-18}) \times E \times \left(\frac{n_B}{n_\gamma}\right) \text{ e cm} \quad (3)$$

where E is now the total unexpected entropy generated by hiccups in the expanding Universe at $T < m_X \approx 10^{15}$ GeV. Setting $(n_B/n_\gamma) \geq 2 \times 10^{-10}$ and using the latest experimental upper bound²⁷ on d_n of 6×10^{-25} e cm, we find from equation (3) that

$$E \leq 1.2 \times 10^3 \quad (4)$$

This restriction on entropy generation after baryosynthesis questions the idea of breaking the Weinberg-Salam $SU(2) \times U(1)$ symmetry by radiative corrections, because this would

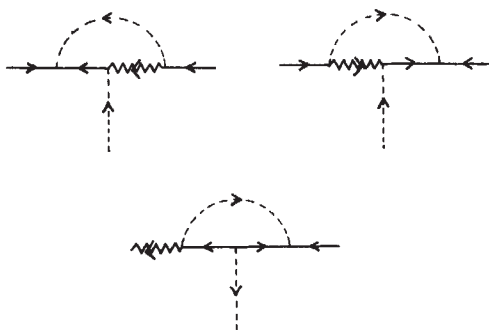


Fig. 1 Third-order renormalizations of the Higgs-fermion-fermion vertex which change θ in an $SU(5)$ model with two $\underline{5}$ representations of Higgs fields. The dashed lines are Higgses, the solid lines $\underline{10}$ fermions and zigzags $\underline{5}$ fermions.

generate^{28,29} an extra entropy factor $E \sim 10^5$ when $T \approx 1$ GeV. Another application of the bound of equation (4) combined with the estimate of equation (2) is to restrict primordial shear to $\Sigma(m_X) \leq 0(10^{18})$. This bound on $\Sigma(m_X)$ is similar to that quoted by Bond, Kolb and Silk²⁰, but does not depend on an estimate of the primordial CP-violating parameter responsible for baryon generation.

Our bound (3) is actually rather conservative, because d_n is considerably larger in all the GUTs we have studied. For example, in an $SU(5)$ model with just one $\underline{5}$ representation of Higgs fields, d_n is larger³⁰ by a factor of $0((\alpha_H/\pi)^{-2}) \equiv 0((\alpha_m^2/\pi m_W^2)^{-2})$ than the lower bound (3). In fact, this model has difficulties in getting a large enough n_B/n_γ and the conventional three-generation version gives too small a ratio^{7,31,32}. Going to four generations might seem^{33,34} to give a large enough CP violation if the fourth generation masses were sufficiently large, but these masses destabilize the Higgs potential³⁵⁻³⁷ if they are too large. Furthermore, one should not neglect^{33,34} the evolution of quark mass parameters between present energies and 10^{15} GeV. One expects^{38,39} them to diminish by a factor of 3 in this energy range, thus diminishing the attainable CP violation by a factor of $3^6 \approx 0(10^3)$. In any case, the phenomenological ratio of the b quark and τ lepton masses argues against^{38,39} increasing the number of generations beyond three. In an $SU(5)$ model with two $\underline{5}$ representations of Higgs fields^{9,40,41}, the leading contribution to d_n is expected to be $0((\alpha_H/\pi)^{-1}) = 0((\alpha_m^2/\pi m_W^2)^{-1})$ larger than the bound (2), and comes through renormalization of the QCD θ vacuum parameter as shown in Fig. 1. However, it is possible neither to make a quantitative estimate nor to exclude a cancellation in this order because of the arbitrariness of the Yukawa coupling parameters. In an $SU(5)$ model with a $\underline{5}$ and a $\underline{45}$ of Higgs fields^{34,42,43}, the leading contribution to d_n is $0((\alpha/\pi)^{-1})$ larger than the bound (2), provided one takes the plausible point of view that most of the baryon asymmetry is generated by Higgs boson decay. In fact, the observed m_b/m_τ ratio suggests that the masses of these heavy fermions do not receive a substantial contribution from a $\underline{45}$ of Higgs, although there may be a role for it in lighter fermion masses^{34,42,43}. Both these $SU(5)$ models with two Higgs representations contributing masses to fermions suggest $\theta \sim (\alpha/\pi) m_a^2/m_W^2$ and hence values for d_n very close to the experimental upper limit. Thus the need to complicate GUTs to get a large enough baryon asymmetry may have other observable consequences.

On the basis of the three models discussed above, we are tempted to strengthen our previous¹⁰ bound (3) to

$$d_n \geq 0\left(\left(\frac{\alpha}{\pi}\right) \times 2.5 \times 10^{-18}\right) \times E \times \left(\frac{n_B}{n_\gamma}\right) \text{ e cm} \\ \geq 0(6 \times 10^{-26}) E \text{ e cm} \quad (5)$$

so that the entropy factor $1 \leq E \leq 0(10)$; but this may be premature until we know which is the right GUT. Even if we did know our GUT, the calculational uncertainties are unlikely ever to exclude a factor of 10 extra entropy generation. The improved bound on E in equation (5) translates into a restriction $\Sigma(m_X) \leq 0(10^{12})$ which is likely to be the best bound on primordial shear that one could ever obtain in this way. As the minimal shear inhomogeneity $\delta\Sigma(m_X)$ that one needs is $0(10^{-3})$, this argument is never likely to exclude the reconciliation of GUTs and galaxy formation from isothermal perturbations²⁰.

If a neutron electric dipole moment d_n were to show up at a level appreciably higher than that expected if $E = 0(1)$, it would suggest that the Universe may have had an irregular expansion after baryon generation. As an example, we have pointed out here that d_n tells us something significant about primordial shear. The neutron electric dipole moment is a sensitive cosmological seismometer.

We thank J. Silk for stimulating conversations.

Received 27 April; accepted 15 July 1981.

1. Sakharov, A. Zh. eksp. teor. Fiz. **5**, 32 (1967); **76**, 1172 (1979).
2. Kuzmin, V. A. Zh. eksp. teor. Fiz. **13**, 335 (1970).

3. Ignatiev, A. Yu., Krasnikov, N. V., Kuzmin, V. A. & Tavkhelidze, A. N. *Phys. Lett.* **76B**, 436 (1978).
4. Yoshimura, M. *Phys. Rev. Lett.* **41**, 381 (1978); **E42**, 746 (1979); *Phys. Lett.* **88B**, 294 (1979).
5. Dimopoulos, S. & Susskind, L. *Phys. Rev. D* **18**, 4500 (1978); *Phys. Lett.* **81B**, 416 (1979).
6. Toussaint, D., Treiman, S. B., Wilczek, F. & Zee, A. *Phys. Rev. D* **19**, 1036 (1979).
7. Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. *Phys. Lett.* **80B**, 360 (1979); **E82B**, 464 (1979).
8. Weinberg, S. *Phys. Rev. Lett.* **42**, 850 (1979).
9. Nanopoulos, D. V. & Weinberg, S. *Phys. Rev. D* **20**, 2484 (1979).
10. Ellis, J., Gaillard, M. K., Nanopoulos, D. V. & Rudaz, S. *Phys. Lett.* **99B**, 101 (1980).
11. Turner, M. S. & Schramm, D. N. *Nature* **279**, 303 (1979).
12. Press, W. *Phys. Scripta* **21**, 702 (1980).
13. Silk, J. *The Cosmic Microwave Background Radiation* (University of California, Berkeley preprint, 1980).
14. Peebles, P. J. E. *The Large-scale Structure of the Universe* (Princeton University Press, New Jersey, 1980).
15. Fabbri, R., Guidi, I., Melchiorri, F. & Natale, V. *Phys. Rev. Lett.* **44**, 1563 (1980).
16. Boughn, S. P., Cheng, E. S. & Wilkinson, D. T. *Astrophys. J. Lett.* **243**, 113 (1981).
17. Silk, J. & Wilson, M. L. *Astrophys. J. Lett.* **244**, 37 (1981).
18. Smoot, G. F., Gorenstein, M. V. & Muller, R. A. *Phys. Rev. Lett.* **34**, 898 (1977).
19. Barrow, J. D. & Turner, M. S. *Nature* **291**, 469 (1981).
20. Bond, J. R., Kolb, E. W. & Silk, J. *The Generation of Isothermal Perturbations in the Very Early Universe* (University of California, Berkeley preprint, 1981).
21. Guth, A. H. *Phys. Rev. D* **23**, 347 (1981).
22. Bais, F. A. & Rudaz, S. *Nucl. Phys. B* **170**, (FS1), 507 (1980).
23. Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. *Phys. Lett.* **90B**, 253 (1980).
24. Schramm, D. N. Preprint EFT-81-03 (Enrico Fermi Institute, 1981).
25. Barrow, J. D. *Mon. Not. R. astr. Soc.* **175**, 359 (1976).
26. Olson, D. W. & Silk, J. *Astrophys. J.* **226**, 50 (1978).
27. Altarev, I. S. *et al.* Preprint 636 (Leningrad Nuclear Physics Institute, 1981).
28. Witten, E. *Nucl. Phys. B* **177**, 477 (1981).
29. Guth, A. H. & Weinberg, E. J. *Phys. Rev. Lett.* **45**, 1131 (1980).
30. Ellis, J. & Gaillard, M. K. *Nucl. Phys. B* **150**, 141 (1979).
31. Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. *Unification of the Fundamental Particle Interactions* (eds Ferrara, S., Ellis, J. & van Nieuwenhuizen, P.) 461 (Plenum, New York, 1980).
32. Barr, S., Segre, G. & Weldon, H. A. *Phys. Rev. D* **20**, 2494 (1979).
33. Segre, G. & Turner, M. S. *Phys. Lett.* **99B**, 399 (1981).
34. Harvey, J. A., Kolb, E. W., Reiss, D. B. & Wolfram, S. Preprint CALT 68-815 (Caltech, 1981).
35. Krasnikov, N. V. *Yad. Fiz.* **28**, 549 (1978).
36. Hung, P. Q. *Phys. Rev. Lett.* **42**, 873 (1979).
37. Politzer, H. D. & Wolfram, S. *Phys. Lett.* **82B**, 242 (1979); **E83B**, 421 (1979).
38. Buras, A. J., Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. *Nucl. Phys. B* **135**, 66 (1978).
39. Nanopoulos, D. V. & Ross, D. A. *Nucl. Phys. B* **157**, 273 (1979).
40. Ignatiev, V. A., Kuzmin, V. A. & Shaposhnikov, M. E. *Zh. eksp. teor. Fiz.* **30**, 726 (1979).
41. Yildiz, A. & Cox, P. H. *Phys. Rev. D* **21**, 906 (1980).
42. Georgi, H. & Jarlskog, C. *Phys. Lett.* **86B**, 297 (1979).
43. Georgi, H. & Nanopoulos, D. V. *Nucl. Phys. B* **159**, 16 (1979).

Emission line regions and stellar associations in extended extragalactic radio sources

D. S. De Young

Kitt Peak National Observatory, Tucson, Arizona 85726, USA

Recent observations of two radio galaxies, 3C277.3 and Centaurus A, reveal the coexistence of emission line filaments and the extended radio-emitting regions. Coincident associations of young stars are also found in Cen A. A model is proposed here that can account for the presence of both emission line regions and young stars as arising from the entrainment of interstellar matter by the outflowing gas which powers the radio source. Late evolution of the young stars may provide a significant fraction of the *in situ* energy injection required for the radio sources.

It has long been known that emission line filaments and early-type stars lie adjacent to the inner north lobe of the nearest radio galaxy, NGC5128 (refs 1, 2). Recent observations of this galaxy by Graham and Price³ confirm this result and reveal additional associations of young stars in and along the boundary of the radio source. These stars lie as much as ~40 kpc from the nucleus of the galaxy and thus cannot have been ejected from the nucleus unless the ejection velocity is ~c, which is extremely unlikely for a star of mass 10–20 M_⊙. Hence these associations of young stars must have been formed *in situ*.

A similar situation has been found in the distant radio galaxy Coma A (3C277.3) by Miley *et al.*⁴ and Bridle *et al.*⁵. Here emission line regions have been seen to coincide with the radio-emitting regions both to the north and south of the galaxy together with continuum emission coincident with radio hot spots in the south. The distance of this galaxy (~300 Mpc) precludes detection of young stellar associations unless they are

very large and compact; quantitative limits will be provided later. These observations reveal the presence of cool (~10⁴ K), metal-enriched gas in coexistence with the population of relativistic particles and magnetic field needed to power the radio source at distances of ~60 kpc from the parent galaxy.

It is suggested here that this phenomenon is associated with a certain class of radio source, namely the Fanaroff and Riley 'Class I' objects which show a relaxed, diffuse morphology with bright spots often occurring near the parent galaxy. Observations of these sources suggest^{6,7} that the relative motion of energetic material ejected from the galaxy is sonic or subsonic with respect to the surrounding medium. In the context of beam models such flows are probably stable but will develop a boundary layer that can entrain and impart momentum to gas from the surrounding medium. The role of this entrained gas in explaining the observed emission line filaments and young stars at large distances from active radio galaxies is now explored.

First the amount of mass that may be entrained by a nearly sonic jet as it passes through the interstellar or circumgalactic medium is estimated. For Mach numbers ≲ 2, the development of the turbulent boundary layer surrounding sonic and supersonic jets has been found experimentally to be very similar to that of subsonic flow⁸. The entrainment for such flows is determined empirically to be⁹

$$\Delta M \approx \frac{1}{4} \pi^2 \rho_0 R_c^2 R_j \quad (1)$$

where ΔM is the external mass entrained by a large-scale eddy, ρ_0 is the density of the ambient medium, and R_j and R_c are the respective radii of the jet and the large-scale eddies in the turbulent boundary layer. The entrainment rate is given by ΔM times the production rate of large-scale eddies, or $M \approx \Delta M u_c / R_c$, where u_c is the convection speed. This result is consistent with the rate derived by Brown and Roshko⁹. Setting $R_c = \delta R_j$ ($\delta < 1$) and $u_c \sim$ sound speed in the boundary layer gives

$$M \approx 60 R_{j,\text{kpc}}^2 \delta n_0 c_s M_\odot \text{ yr}^{-1} \quad (2)$$

where n_0 is the ambient number density and c_s is the sound speed in units of 10⁸ cm s⁻¹. The experimental results not only provide the entrainment rate but also show that most of the entrainment occurs during the formation of the large scale eddies. Subsequent evolution of the boundary layer primarily involves a mixing of the entrained material on successively finer scales; hence equation (2) is not proportional to the jet length.

If the jet is moving supersonically with respect to the surrounding medium, gas entrained in the turbulent boundary layer will have been shock preheated to a temperature that corresponds to a sound speed comparable with the jet velocity, that is $c_s \sim 1$. The cooling time for this gas must be less than or of the order of the travel time of the beam that forms the radio source to account for the observed filaments and possible star formation. The post-shock temperature¹⁰ is given by $T_s = 1.4 \times 10^5 v_{s,7}^2$ K, where $v_{s,7}$ is the shock speed in units of 10⁷ cm s⁻¹. For $v_{s,7} = 10$, $T_s \approx 10^7$ K, and for a gas with solar abundances the cooling time¹¹ is $t_c = 3 \times 10^2 T^{1.6} / n$ s for $3 \times 10^5 \text{ K} \leq T \leq 4 \times 10^7 \text{ K}$, where n is the number density. If $T = T_s \approx 10^7 \text{ K}$, $t_c = 1.5 \times 10^6 / n$ yr. These cooling times apply locally only after the initial passage of the jet; at later times near the base of the jet the ambient gas will have had time to cool to $\leq 10^5 \text{ K}$ and can be entrained at this temperature. Capture by the fast-moving jet can reheat this gas to no more than $\sim T_s$, thus the cooling times given above are upper limits.

A likely site for entrainment of metal-enriched gas is the interstellar medium of the galaxy through which the jet passes. The characteristics of this medium for early-type galaxies producing radio sources are uncertain, but X-ray data¹² for M87 are consistent with average number densities between 10⁻² and 10⁻¹ cm⁻³ and temperatures of ~10⁷ K. Similar values for the interstellar density were found by Jones and Owen¹³ to be consistent with the distortion of the radio-emitting plasma in NGC1265. This provides cooling times of 10⁷–10⁸ yr, during

which the entrained material moves a distance

$$R = 10.5 v_8 t_7 \text{ kpc} \quad (3)$$

with the jet velocity $v_j = 10^8 v_8 \text{ cm s}^{-1}$ and $t = 10^7 t_7 \text{ yr}$. Thus, when the enriched gas has cooled to $\sim 10^4 \text{ K}$ and produces optical emission lines, it will have travelled 10–100 kpc from the parent galaxy, which agrees well with the observations. These cooling times will be reduced if the entrainment process produces local density enhancements. Further entrainment could occur if the jet passes through an extended circumgalactic halo with a significant metal-enriched component produced by a galactic wind¹⁴.

The total mass entrained is obtained from equation (2). Using $n_0 = 0.1$ – 0.01 and $\delta = 0.1$ (an underestimation when contrasted with laboratory experiments¹⁵) gives a total mass of 6×10^7 – $6 \times 10^6 M_\odot$ entrained over 10^8 yr . The total mass observed in the emitting line regions in Coma A varies from 10^5 to $10^9 M_\odot$, depending on the filling factor, with the most likely value somewhere near the logarithmic mean of these limits.

Thus, on the basis of purely hydrodynamic considerations, the observed emission line regions associated with extended radio sources may be accounted for by the entrainment and cooling of metal-enriched gas along the boundary of the jet or beam that gives rise to the radio source. This picture has the advantage of not requiring the beam itself to be contaminated *ab initio* with large amounts of cool enriched plasma that would impair the efficiency of its formation. It also predicts that many other Class I radio sources will show evidence of emission line regions coincident with the radio source. The process has another implication in terms of radio source energetics, namely the possibility of star formation occurring in the entrained and co-moving emission line regions.

If some fraction of the gas in the emission line regions is to cool further, collapse, and form stars, these fragments must be shielded from any ionizing UV radiation. Miley *et al.*⁴ estimate the flux of ionizing photons emanating from the bright optical knots in Coma A to be $\phi \sim 10^9 \text{ cm}^{-2} \text{ s}^{-1}$ at 1 kpc from the knot. The shielding depth is then¹⁶

$$l_s = \frac{\phi}{\alpha n^2} \quad (4)$$

where α is the recombination coefficient. The value of n in the emission line region is uncertain because of the filling factor; Miley *et al.*⁴ provide limits of 0.3 – 10^4 cm^{-3} . This gives values of l_s from 10^{13} cm to $\sim 1 \text{ kpc}$, with correspondingly smaller values further from the bright knots. The metal-enriched gas which is shielded can cool to $T \sim 10^2 \text{ K}$, and the Jeans mass and corresponding length,

$$M_J = \frac{40 T^{3/2}}{n^{1/2}} M_\odot; \quad \lambda_J = 15 \left(\frac{T}{n} \right)^{1/2} \text{ pc} \quad (5)$$

lie in the range 6.7×10^4 – $400 M_\odot$ and 250 – 1.5 pc on using the above limits for n . These values are overly large, as the number density in the collapsing fragment at $T \sim 10^2 \text{ K}$ will be greater than that observed in the parent emission line region at 10^4 K . If the collapse is isobaric, $n \propto T^{-1}$, and the collapse time $t_c \sim (4\pi G \rho)^{-1/2}$ is 10^5 – 10^7 yr .

The size of the emission line regions in Coma A is $\geq 1 \text{ kpc}$ near the continuum knots and could be much larger than this in the northern portions of the radio source. Thus the values of l_s , M_J , λ_J and t_c are all consistent with the observed quantities and suggest that star formation could be taking place in the environs of the extended radio source. For Cen A, emission line filaments are a few hundred parsecs in size or larger, with masses $\sim 10^6 M_\odot$ and electron densities of $\sim 0.2 \text{ cm}^{-3}$. Graham and Price³ argue that shock-wave excitation dominates in this object, so that the constraints imposed by l_s are inapplicable. The resulting values of M_J and λ_J again imply star formation along the northern radio source, and indeed a minimum of $10^3 M_\odot$ in young ($10 M_\odot$) stars is observed in this region. (The distribution of these stars provides no constraints on R_j since they are imbedded primarily along one side of the radio lobe.)

The details of star formation in a cooling turbulent boundary layer are obscure. If, however, a 'Salpeter' mass function ($dN/dM \propto M^{-2.35}$) is used with upper and lower cutoffs of 30 and $0.1 M_\odot$, $\sim 3\%$ of all stars formed will have $M \geq 10 M_\odot$, or in terms of the total mass in stars, $N(M \geq 10) = 1.3 \times 10^{-2} (M_{\text{tot}}/M_\odot)$. These massive stars are of interest because of their short lifetimes, $\tau_{\text{ms}} \sim 10^7 \text{ yr}$. Note that these stars are formed from entrained mass in the beam boundary layer and thus share the outward motion of that gas at the rate of $\sim 10 \text{ kpc}$ every 10^7 yr . At least 100 such stars are seen in Cen A; in more distant galaxies they would be visible if $m_B \leq 23 \text{ arc s}^{-2}$, which implies

$$L_{\text{max}} = 2 \times 10^6 D_{100}^2 L_\odot \quad (6)$$

where D_{100} is the distance in 100 Mpc units. For Coma A, this limit is reached at 1,000 stars of $10 M_\odot$ in a single cluster. The fraction of entrained mass that forms stars is unknown; if $\sim 10\%$ efficiency is assumed, equation (1) implies 10^6 – $10^7 M_\odot$ in stars formed continuously over 10^8 yr , of which 10^4 – 10^5 could have masses $\geq 10 M_\odot$, with 10^3 – 10^4 present at any one time as $\tau_{\text{ms}} \sim 10^7 \text{ yr}$.

At the end of their lifetime such massive, co-moving stars will explode, yielding $\sim 10^{51}$ – 10^{52} erg per event and leaving a remnant that can produce $\sim 10^{38} \text{ erg s}^{-1}$. The summation of such events has a significant effect on the total energy budget of the radio source. After the first 10^7 yr , the remnants alone can inject $10^{41} \text{ erg s}^{-1}$ and 10 times this value at the 'current' age of $\sim 10^8 \text{ yr}$. The supernova events can supply a comparable average energy input. It is generally recognized⁶ that 'noisy' Class I radio sources with sonic or subsonic outflow require in many cases *in situ* particle re-energization along the beam as well as in the more remote regions of the radio source. Massive stars and their remnants, moving outward with velocities comparable with the beam velocity, can provide sources of energy for reacceleration of the electrons in the radio source. This could occur either through direct interaction with the supernova remnant or less efficiently via kinetic energy injection and subsequent stochastic acceleration from the supernova event itself. It should be noted that the evolution of these massive stars does not produce the radio source, but instead provides a means of reaccelerating a pre-existing power law distribution of particles in a very low density plasma. Energy injection of $10^{42} \text{ erg s}^{-1}$ is comparable with that required for many Class I radio sources. This number could be much larger given different values for the slope of the mass function, the efficiency of star formation and entrainment, and the density of the surrounding medium.

Would such supernovae in radio galaxies be seen? Complete, systematic surveys for supernovae extend only to small distances¹⁷ ($\leq 30 \text{ Mpc}$) beyond which most radio galaxies lie, and in addition no search has been conducted for supernovae lying well outside the stellar component of radio galaxies. Such supernovae could be observed in distant ($> 100 \text{ Mpc}$) radio sources if a large-aperture telescope were used, but no efforts have yet been made in this direction.

I thank Dr H. Butcher for helpful discussions. The Kitt Peak National Observatory is operated by the Association of Universities for Research in Astronomy, Inc., under NSF contract.

Received 24 April; accepted 24 June 1981.

1. Blanco, V. M., Graham, J. A., Lasker, B. M. & Osmer, P. S. *Astrophys. J. Lett.* **198**, L63–L64 (1975).
2. Osmer, P. S. *Astrophys. J. Lett.* **226**, L79–L82 (1978).
3. Graham, J. A. & Price, R. M. *Astrophys. J.* (in the press).
4. Miley, G. K., Heckman, T. M., Butcher, H. R. & van Breugel, W. J. M. *Astrophys. J. Lett.* (in the press).
5. Bridle, A. H., Fomalont, E. B., Palimaka, J. J. & Willis, A. G. *Astrophys. J.* (in the press).
6. Miley, G. A. *Rev. Astr. Astrophys.* **18**, 165–218 (1980).
7. van Groningen, E., Miley, G. K. & Norman, C. A. *Astr. Astrophys.* **90**, L7–L9 (1980).
8. Bradshaw, P. A. *Rev. Fluid Mech.* **9**, 33–54 (1977).
9. Brown, G. L. & Roshko, A. *J. Fluid Mech.* **64**, 775–816 (1974).
10. McKee, C. F. & Hollenbach, D. J. A. *Rev. Astr. Astrophys.* **18**, 219–262 (1980).
11. McKee, C. F. & Cowie, L. L. *Astrophys. J.* **215**, 213–225 (1977).
12. Fabricant, D., Lecar, M. & Gorenstein, P. *Astrophys. J.* **241**, 552–560 (1980).
13. Jones, T. W. & Owen, F. N. *Astrophys. J.* **234**, 818–824 (1979).
14. Mathews, W. G. & Baker, J. C. *Astrophys. J.* **170**, 241–259 (1971).
15. Hinze, J. O. *Turbulence* 420–439 (McGraw-Hill, New York, 1959).
16. Felten, J. E. & Bergeron, J. *Astrophys. Lett.* **4**, 155–157 (1969).
17. Tammann, G. A. *Mem. Soc. Astr. Ital.* **49**, 315–329 (1978).

Formation of organic molecules on Titan

Louis A. Capone*, Sheo S. Prasad†, Wesley T. Huntress†, Robert C. Whitten‡, John Dubach§ & Krishnamurthy Santhanam||

*Department of Meteorology, San Jose State University, San Jose, California 95192, USA

†Jet Propulsion Laboratory, Pasadena, California 91103, USA

‡Space Science Division, NASA-Ames Research Center, Moffett Field, California 94035, USA

§Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA

||Informatics, Inc., Palo Alto, California 94303, USA

The unique atmospheric environment on Titan has stimulated great interest in its organic chemistry. Recently¹ we proposed that simple organic-nitrogen compounds such as HCN could be efficiently formed by cosmic ray bombardment of a nitrogen-containing atmosphere on Titan. Voyager I has now verified that molecular nitrogen is indeed the major constituent on Titan^{2,4} and that HCN is also present². Based on these new data, we now propose that even more complex organic-nitrogen molecules such as ethyl cyanide ($\text{CH}_3\text{CH}_2\text{CN}$), vinyl cyanide (CH_2CHCN), and cyanoacetylene (HCCCN) may be formed efficiently in the lower atmosphere of Titan, where lower temperatures and higher densities will ensure the efficacy of three-body ion-molecule association reactions. Interestingly, these compounds have been found in several dark interstellar clouds^{5,6}, thus the chemistry suggested here is analogous to that proposed as an explanation of interstellar cyanopolyynes⁷. The only difference is the role played by three-body association reactions in the dense lower atmosphere of Titan. The mechanism proposed here rests upon a firmer experimental foundation than the analogous radiative association reactions in interstellar clouds.

The model atmosphere we adopted is consistent with the preliminary interpretation of Voyager 1 data in that it contains nitrogen and methane in the ratio 98.8:0.8% as well as trace amounts of C_2H_2 , H_2 , C_2H_4 and C_2H_6 formed from methane photolysis. The temperature and pressure in the region of maximum cosmic ray-activated chemistry² are ~ 150 – 160 K and 20 mbar. The detailed calculation of the cosmic ray-induced ionization rates has been discussed elsewhere^{1,8–11}. These calculations took into account the production, decay and absorption of secondary cosmic-ray cascade particles such as protons, charged and neutral pions, muons and electrons. The corresponding maximum electron concentration is $\sim 2 \times 10^3 \text{ cm}^{-3}$.

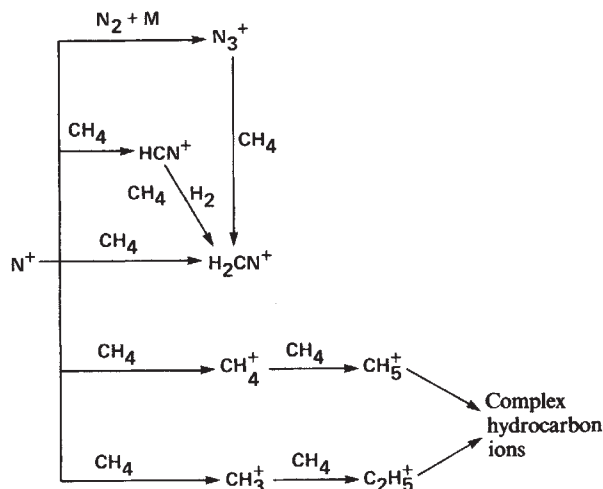


Fig. 1 Schematic diagram of N^+ ion-molecule reactions with CH_4 and N_2 .

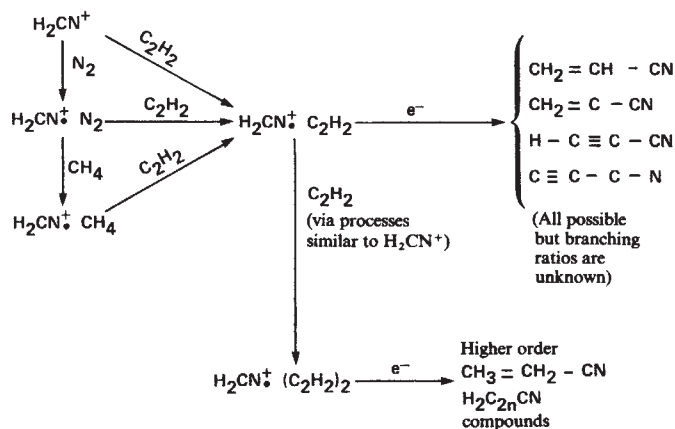


Fig. 2 The synthesis of complex organic-nitrogen ions resulting from the production of HCNH^+ in Titan's atmosphere.

The ionization of CH_4 and the reactions of N_2^+ ions with CH_4 yield only hydrocarbon ions^{12–14}. Reaction of N^+ (formed by the dissociative ionization of N_2) with CH_4 can, however, yield HCNH^+ ions (Fig. 1). The lifetimes of these ions against dissociative recombination and proton transfer to NH_3 (if present at a level of 0.01 parts per 10^9 by volume) are $\sim 10^3$ s. Hence this ion starts a chain of interesting three-body association reactions and bimolecular switching reactions, shown in Fig. 2, which may involve N_2 , CH_4 , C_2H_2 and C_2H_4 . As N_2 and CH_4 are far more abundant than C_2H_2 , C_2H_4 and electrons, clustering of an ion with N_2 and CH_4 will overwhelm other reactions such as dissociative recombination, insertion of C_2H_2 , or the displacement of N_2 by C_2H_2 . The terminal cluster ion can either recombine with electrons or add an additional C_2H_2 or C_2H_4 by three-body reactions to produce a larger carbon chain molecular ion. The fraction which recombines leads to the formation of organic-nitrogen compounds. The following is an example of these ideas.

The formation of the $\text{CH}_2\text{CHCNH}^+$ ion by the three-body association reaction



has been found in laboratory experiments¹⁵ to have a rate coefficient of $\sim 10^{-28} \text{ cm}^6 \text{ s}^{-1}$ at 300 K; the clustering of HCNH^+ with N_2 or CH_4 leading to $\text{HCNH}^+ \cdot \text{N}_2$ or $\text{HCNH}^+ \cdot \text{CH}_4$, has not been investigated experimentally. Formation of a weakly bound $\text{HCNH}^+ \cdot \text{N}_2$ cluster ion could still lead to $\text{CH}_2\text{CHCNH}^+$ formation through the switching reactions



as shown in Fig. 2. The fraction of $\text{HCNH}^+ \cdot \text{C}_2\text{H}_2$ ions which recombine with electrons may lead to the formation of ethyl cyanide, vinyl cyanide and cyanoacetylene (see Fig. 2), provided the cluster ion has a suitable structure. The fact that saturated hydrocarbons are common in nature suggests that this assumption is not unfounded. The remaining fraction of $\text{HCNH}^+ \cdot \text{C}_2\text{H}_2$ cluster ions might undergo the same series of clustering and switching reactions which produced $\text{HCNH}^+ \cdot \text{C}_2\text{H}_2$ from the initial H_2CN^+ ion. The end product would be $\text{H}_2\text{CN}^+ \cdot (\text{C}_2\text{H}_2)_2$ ions whose dissociative recombination would yield higher order $\text{H}_x\text{C}_{2n}\text{CN}$ compounds (see Fig. 2). Free electrons would be present due to the rarity of electron-attaching species such as CH_3 , H and NH_2 .

The total production rate of all complex organic-nitrogen compounds is determined by the competition between the dissociative recombination of $\text{H}_2\text{CN}^+ \cdot \text{CH}_4$ and its switching reaction with C_2H_2 . In either case, a significant production rate for complex organic-nitrogen compounds is expected on Titan. The total production rate shown in Fig. 3 is based on assumed values of the rate coefficient for switching reactions (2) and (3) of $10^{-10} \text{ cm}^3 \text{ s}^{-1}$ and for three-body association reaction (1) of

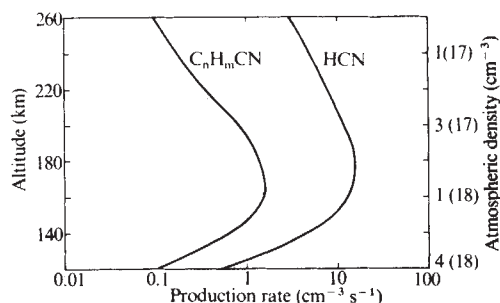


Fig. 3 Predicted production rates of organic nitriles ($C_n H_m CN$) shown as a function of altitude and atmospheric density. Also shown is the calculated production rate of hydrogen cyanide (HCN), assuming the neutral model described in the text.

$10^{-27} \text{ cm}^6 \text{ s}^{-1}$ at a temperature of $\sim 150 \text{ K}$; the three-body coefficient is further assumed to saturate at $10^{-10} \text{ cm}^3 \text{ s}^{-1}$. The rate coefficient of $10^{-10} \text{ cm}^3 \text{ s}^{-1}$ for the ion-molecule switching reactions (2) and (3) is a reasonable estimate for that type of reaction. For the three-body ion cluster reactions of H_2CN^+ with C_2H_2 , CH_4 , or N_2 , we have adopted McEwan's measured rate coefficient ($10^{-28} \text{ cm}^6 \text{ s}^{-1}$) for reaction (1) with an approximate T^{-3} -dependence, which is appropriate for such reactions¹⁶. Figure 3 shows substantial production rates which were, in fact, found to be relatively insensitive to uncertainties in the rate coefficients for reactions (1)–(3).

Given the production rate of Fig. 3, the equilibrium concentrations of the individual organic-nitrogen compounds will be determined by the branching ratio in reactions which form vinyl cyanide and cyanoacetylene, and by the loss rates of the individual species. At low temperature, neutral chemical reactions are not expected to constitute important mechanisms of loss. Ion-molecule reactions could lead to a loss rate of no greater than 10^{-7} s^{-1} . Also, photodissociation into reactive radicals is rather inhibited due to the optically thick layer of aerosol particles (in the UV and visible ranges) which could act as a protective shield against dissociative UV flux. Thus, significant equilibrium concentrations of these species, perhaps several parts per billion, are expected on Titan. We have used published data¹⁷ to make some simple calculations of the condensational temperatures corresponding to the maximum amounts of complex organics (that is, HCN and CH_3CH_2CN) expected on Titan. Our results show that these species should not freeze out in the 10–40 mbar region.

In view of the above discussions and the Voyager 1 observations of atmospheric composition, it now seems evident that Titan's atmosphere may provide a rich environment for the synthesis of complex organic molecules. Additional laboratory work on the cluster ion involving N_2 , CH_4 and C_2H_2 is needed to elucidate this possibility.

Note added in proof: Since acceptance of this paper, cyanoacetylene (HCCCN) has been detected on Titan at mixing ratios of $\sim 10^{-8}$ to 10^{-8} (V. Kunde, personal communication).

Solitary waves in the lower atmosphere

D. R. Christie & K. J. Muirhead

Research School of Earth Sciences, Institute of Advanced Studies, Australian National University, Canberra, ACT 2600, Australia

R. H. Clarke

Department of Meteorology, School of Earth Sciences, University of Melbourne, Parkville, Victoria 3052, Australia

Solitary waves are of intense interest in the physical and mathematical sciences¹. These nonlinear waves often seem to have a primary role in the asymptotic description of propagating disturbances in inland lakes and coastal waters^{2,4}, in the thermocline of the open sea⁵ and in the lower atmosphere^{6,7}. We present here new acoustic sounder observations of complex tropospheric solitary-wave disturbances at Tennant Creek in the arid interior of Australia (Fig. 1), a description and interpretation of a new type of visible wave phenomena over northern Australia which appear as thin propagating cumulus cloud lines, and a discussion of observations at Burketown on the Gulf of Carpentaria of a new class of low-altitude propagating solitary-wave roll clouds which originate to the south. These observations, when correlated with observations at Tennant Creek, indicate that solitary-wave-generating disturbances in the form of internal bores^{8,9} propagate over large distances.

Nonlinear internal waves on a boundary layer waveguide in an infinitely deep fluid are described to first order by the Benjamin-Ono equation^{10,11}. The general properties of the solutions of nonlinear evolution equations of this type are well known^{1,12–22}. Arbitrarily long, but finite, initial disturbances in the form of long waves of elevation or bores of limited spatial extent evolve into a finite number of solitary waves followed by a complex dispersing oscillatory wave train^{14,17,18}. Infinitely long smooth bores or waves of elevation develop undulations along the leading edge which evolve asymptotically into solitary waves²². These characteristic patterns provide a guide for the interpretation of evolving long internal wave disturbances in the atmosphere.

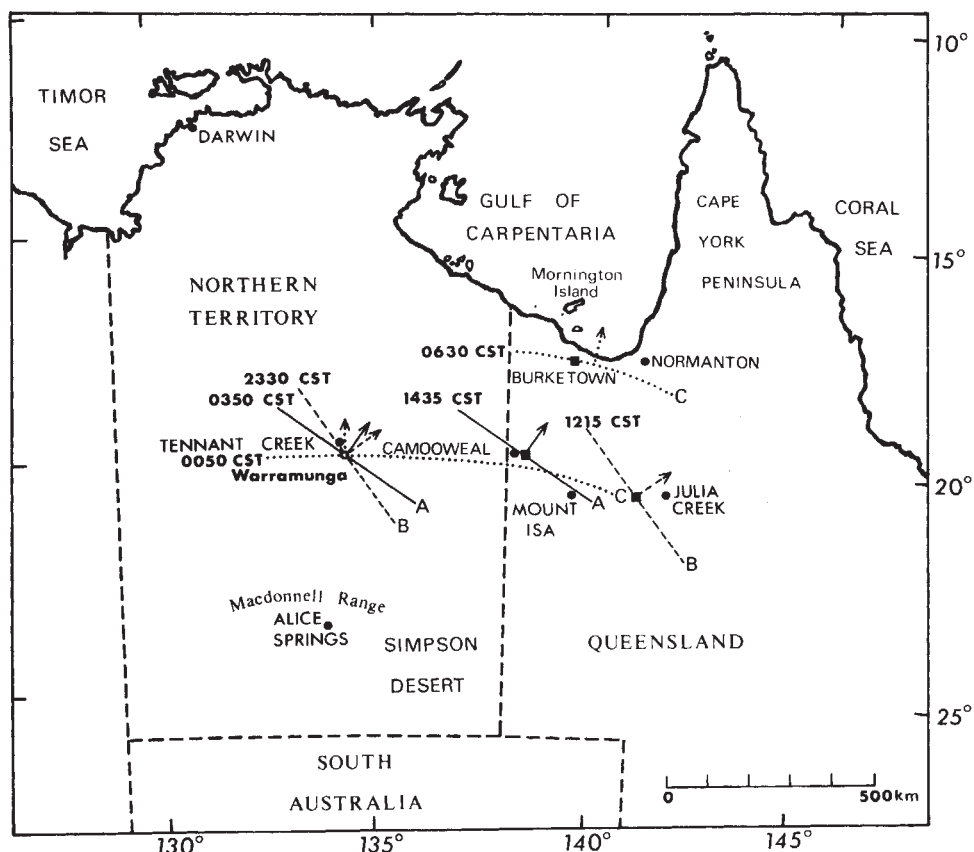
The properties of solitary waves in the lower troposphere have been determined from surface micropressure observations recorded continuously over a 5-yr period on a high-sensitivity infrasonic array⁶ located at Warramunga near Tennant Creek (Fig. 1), and from observations⁷ of solitary waves associated with the Morning Glory phenomenon^{23,24} of the Gulf of Carpentaria. Observations have been made of a wide variety of nonlinear wave phenomena ranging in form from essentially pure asymptotic solitary-wave states to complex evolving solitary-wave-dominated disturbances and undular internal bores which seem to represent early stages in the production of solitary waves. The detailed structure of two extensive solitary-wave-dominated disturbances may be seen in the acoustic facsimile records from Warramunga (Figs 2, 3). These particular events are of only modest amplitude compared with many of the observations recorded on the infrasonic array. In both cases the solitary wave component appears along the leading edge of a more extensive disturbance in the form of a complex internal bore.

A noteworthy feature of internal solitary waves is the closed circulation cell which develops in the interior of large-amplitude solitary waves^{11,17,25}. If sufficient moisture is present solitary waves may appear as horizontal propagating retrograde low-level roll clouds. The cloud formation accompanying the Morning Glory of the Gulf of Carpentaria is a spectacular phenomenon of this type⁷. Robin²⁶ has described a propagating roll cloud family over Spencer Gulf in South Australia which may also be interpreted as a visible manifestation of solitary waves propagating on a marine inversion.

Received 18 May; accepted 30 June 1981.

- Capone, L. A., Dubach, J., Whitten, R. C., Prasad, S. S. & Santhanam, K. *Icarus* **44**, 72–84 (1980).
- Hanel, R. *et al. Science* **212**, 192–200 (1981).
- Tyler, G. L. *et al. Science* **212**, 192–212 (1981).
- Broadfoot, A. L. *et al. Science* **212**, 201–206 (1981).
- Kroto, H. W. *et al. Astrophys. J. Lett.* **219**, L133–L137 (1978).
- Broten, N. W. *et al. Astrophys. J. Lett.* **223**, L105–L107 (1978).
- Mitchell, G. F., Huntress, W. T. & Prasad, S. S. *Astrophys. J.* **233**, 102–108 (1979).
- Dubach, J., Whitten, R. C. & Sims, J. S. *Planet. Space Sci.* **22**, 525–536 (1974).
- Capone, L. A., Whitten, R. C., Dubach, J., Prasad, S. S. & Huntress, W. T. *Icarus* **28**, 367–378 (1976).
- Capone, L. A., Whitten, R. C., Dubach, J. & Prasad, S. S. *Astrophys. J.* **213**, 977–983 (1977).
- Capone, L. A., Dubach, J., Whitten, R. C. & Prasad, S. S. *Icarus* **39**, 433–449 (1979).
- Huntress, W. T. *Astrophys. J. Suppl.* **33**, 495–514 (1977).
- Albritton, D. L. *Atom. Data Nucl. Data Tabl.* **22**, 1–101 (1978).
- Huntress, W. T., McEwan, M. J., Karpas, Z. & Anicich, V. G. *Astrophys. J. Suppl.* **44**, 481 (1980).
- McEwan, M. J. *IAU Symp.* **87**, 299, (1980).
- Smith, D. & Adams, N. G. *Astrophys. J. Lett.* **220**, 87–92 (1978).
- Handbook of Physics and Chemistry*, 43rd edn, p. 2453 (Chemical Rubber Co., Cleveland, 1961).

Fig. 1 Map of northern Australia showing the location (○) of the infrasonic array and acoustic sounder installations at the Warramunga Infrasonic Array near Tennant Creek and the positions (■) and orientations of three observations of unusual low-altitude propagating cloud line events which correlate with events at 03.50 CST (5 October 1979) (A), 23.25 CST (24 July 1979) (B) and 00.50 CST (6 October 1979) (C) detected during the preceding night at Warramunga. The measured propagation speeds and source azimuths determined from the phased array records are: A: $c = 5.6 \text{ m s}^{-1}$, $\theta = 215^\circ$; B: $c = 11.7 \text{ m s}^{-1}$, $\theta = 235^\circ$; C: $c = 9.8 \text{ m s}^{-1}$, $\theta = 180^\circ$.



The structure of the boundary layer over northern Australia is typically characterized²⁷ by a persistent marked elevated inversion at an altitude of several kilometres overlying, during the night time and early daylight hours, an intense surface-based radiation inversion. With the onset of convective activity the nocturnal inversion is eroded away from its base^{28,29} and usually vanishes before noon. It must be expected that relatively small-scale nonlinear wave disturbances associated primarily with the nocturnal inversion will, under the influence of convection, decay and disappear during the day. Larger-amplitude disturbances which involve the elevated inversion layer may be expected to survive for longer periods of time.

The following observations indicate that long nonlinear wave disturbances propagate over substantial distances along these boundary layer waveguides. On 5 October 1979, at 14.35 CST, two closely spaced propagating quasicontinuous thin cumulus cloud lines were observed from the ground in totally clear conditions aligned along the NW-SE direction near Camooweal, 405 km east of Warramunga as shown by line A in Fig. 1. These stable propagating cloud lines, which extended

form horizon to horizon, had a width and depth of some hundreds of metres and a spacing of some kilometres. This observation is correlated in direction and in time, assuming a mean propagation speed of about 6 m s^{-1} , with the nonlinear wave disturbance shown in Fig. 3, observed about 11 h earlier at Warramunga. The disturbance apparently dissolved during the afternoon of 5 October.

The preferred explanation of the two coherent propagating cloud lines observed near Camooweal is that they represent capping clouds created by the lifting of moist air to the condensation level during the passage of solitary waves associated with the leading edge of a dissipating finite-length internal bore. The character of the solitary wave pattern can be expected to evolve substantially as the nonlinear wave disturbance adjusts to the evolving boundary layer density profile²⁰. It is therefore very unlikely that the observed cloud lines correspond directly to any of the solitary waves shown in Fig. 3.

A similar observation of a single, almost continuous, clearly defined, rapidly propagating cumulus cloud line, which again extended to the limits of visibility, was made in otherwise totally

Fig. 2 Monostatic acoustic sounder facsimile record obtained at Warramunga for the period from 03.50 to 09.15 CST on 22 October 1979 which illustrates the detailed morphology of a well defined solitary-wave-dominated complex disturbance which arrived at the array shortly before the transition of the stably stratified nocturnal boundary layer to a convecting daytime boundary layer. The dark portions of the record represent regions of enhanced echo returns from small-scale temperature inhomogeneities. The corresponding micropressure and surface wind records indicate that the major nonlinear wave disturbance in this record persisted for ~2 h beyond the onset at about 05.37 CST. The minor disturbance starting at about 04.50 CST originates from a different direction and seems to be independent of the main event. The main disturbance seems to represent a fairly late stage in the development of a long wave of elevation or internal bore of finite extent into a family of solitary waves. Only the most significant solitary-wave components are marked on the record. The effective wavelength (FWHM) of these waves varies over ~1.4–2.4 km. Sunrise occurred at 06.11 CST. The increasing depth of the developing mixed layer at the surface and the onset of convective plumes can be seen in the record beyond 07.30 CST.

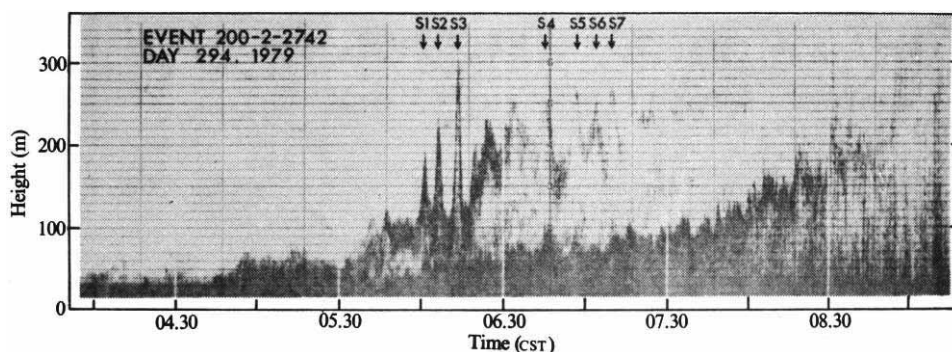
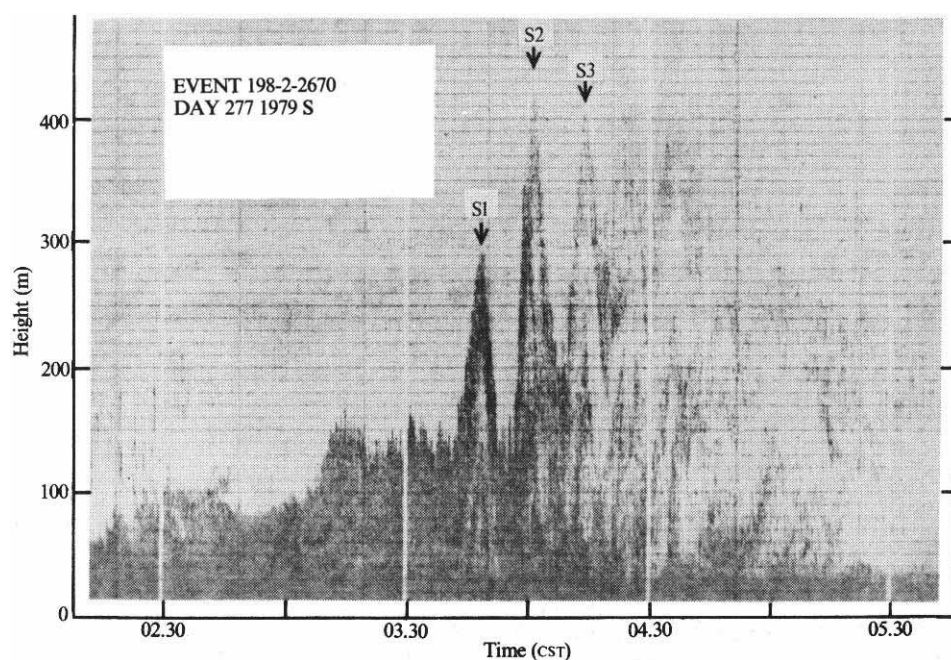


Fig. 3 Acoustic sounder facsimile record obtained on 5 October 1979 at Warramunga which illustrates the structure of a complex nonlinear wave disturbance dominated by a three-component solitary wave family. The corresponding infrasonic array record section shows that the leading solitary wave components are followed by an extensive train of smaller amplitude irregular waves which may represent the subcritical dispersing wave component predicted by nonlinear dispersive wave theory. The meteorological records at Warramunga show that this nonlinear wave disturbance produced only very minor perturbations in the wind field at the surface. These records indicate that the event probably starts at about 03.10 CST. The disturbance is characterized by an increase in surface temperature of 2.5°C which may be attributed to mixing. The fact that the surface winds associated with this event decreased to near zero by 05.00 CST suggests that this disturbance also represents a late stage in the development of an internal bore of limited spatial extent. Note that the effective wavelengths ($\sim 2.9\text{ km}$) of the three solitary waves as indicated by the echo profiles in this record agree well with the infrasonic-array-determined estimates. This solitary wave event is of particular interest because it seems to be well correlated with a visible manifestation of lower tropospheric wave activity observed at 14.35 CST on the same day, 405 km to the east of the array in northern Queensland as shown in Fig. 1.



clear conditions over an essentially featureless plain, 75 km west of Julia Creek (Fig. 1, line B) at 12.15 CST on 25 July 1979. This event was tracked by vehicle for 2 h and it was established that the cloud line was propagating north-east with a speed of about 12 m s^{-1} . During this observational period the structure of the cloud line remained essentially unchanged except for perturbations in the modulation pattern along the line, which may be attributed to the influence of convection. This event can be correlated with an infrasonic array observation at Warramunga at 23.30 CST on 24 July, in the form of a smooth internal bore with an amplitude of 0.8 mbar, with a pressure jump of about 1.5 mbar observed at 07.00 CST on 25 July at Mount Isa airport, and with a marked wind shift (NNE to SW) moving through the anemometer array operated by Mount Isa Mines Ltd. The mean nocturnal speed of 13 m s^{-1} agrees fairly well with the infrasonic array estimate of 11.7 m s^{-1} from 235° . The propagation speed decreased to an average of $\sim 8\text{ m s}^{-1}$ during the morning and then increased to $\sim 12\text{ m s}^{-1}$ during the early afternoon. These variations in speed may be attributed to the influence of topography in the neighbourhood of Mount Isa and to the evolving boundary layer density structure. In this particular event the available wind soundings behind the leading edge of the disturbance show a low-level wind speed component along the direction of motion nearly equal to the propagation speed, suggesting that this disturbance must be associated with a low-level density current. The preferred explanation of this unusual single cloud line is that it represents a capping cloud produced by a large amplitude solitary wave associated with the leading edge of a finite-length internal bore initiated by a density current advancing into the stratified boundary layer. Alternatively, in this case, the propagating cloud line may be due to lift associated with the head of an advancing gravity current.

A different form of solitary-wave-produced cloud formation consisting of two well developed low-level parallel roll clouds, similar in all respects to those associated with the Morning Glory⁷ but originating from the south, was made at Burketown (Fig. 1, line C) at 06.30 CST on 6 October 1979. Visual satellite imagery of these waves is shown in Fig. 4. Remnants of three cloud lines further north with similar orientation were also seen in the visual imagery at 09.30 CST. The cloud lines appeared to propagate with a phase speed of $\sim 15\text{ m s}^{-1}$ from 190° and were coherent with pressure jumps on the barographs at Normanton

(07.00 CST), Mornington Island (08.15 CST) and Mount Isa (0000 CST). Wind data from the Mount Isa Mines Ltd mesoscale array reveal sharp changes from very light southwesterly to fresh southerly winds, gusting above 13 m s^{-1} , unaccompanied by any sign of solitary wave activity. The disturbance was observed at Warramunga at 00.50 CST in the form of an internal bore of

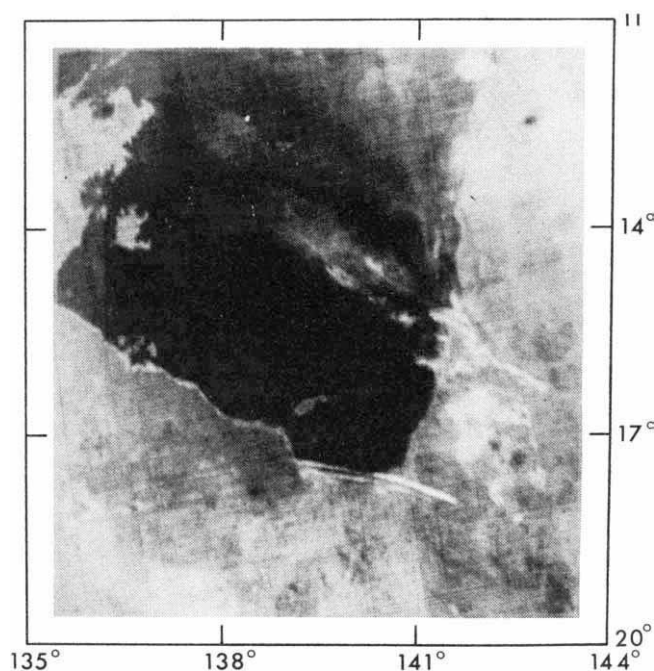


Fig. 4 Visual imagery of the Gulf of Carpentaria area from the DMSP (Defence Meteorological Satellite Program) satellite at 07.07 CST on 6 October 1979, showing the position and orientation of two slightly curved parallel low-altitude roll clouds produced by solitary wave activity in the boundary-layer waveguide. These cloud lines can be correlated with observations of nonlinear wave pressure perturbations and surface wind shear lines during the preceding night at Warramunga and Mount Isa as shown in Fig. 1. The cloud lines are 285 km long, $\sim 13\text{ km}$ apart and $\sim 6\text{ km}$ wide and were observed to propagate towards an azimuth of 10° with a speed of 15 m s^{-1} .

amplitude 0.7 mbar. The array-determined speed (9.8 m s^{-1}) and azimuth (180°) imply a curvature in the wavefront in agreement with the pattern defined by the cloud lines shown in Fig. 4. Again, these observations may be attributed to the passage of a disturbance in the form of an internal bore of finite extent which evolves asymptotically in time into a family of solitary waves.

On all occasions the propagating cloud lines and coherent pressure signatures seem to be closely related to synoptic developments. On the early morning of 5 October a trough lay $\sim 400 \text{ km}$ south-west of Warramunga with a NW-SE orientation. Over the next 36 h this feature, with a following high pressure ridge, moved east across the Australian continent at a speed of about 7 m s^{-1} (at lat. 20°). The observations presented above indicate that on the evening of 5 October, on the south-west side of this trough, evidently as a result of deformation of the nocturnal boundary layer, a propagating internal bore developed in the area south of Tennant Creek and moved NNE at over 15 m s^{-1} on 6 October for at least 700 km. It developed solitary waves, made visible by propagating roll cloud lines in the Gulf area as shown in Fig. 4. The sequence observed on 24–25 July was similar, except that in this case the trough, pressure jump line and wind shear line were coincident throughout their known history, progressing steadily eastward at about 12 m s^{-1} for at least 8 h and, in the case of the trough and shear line, for a further 21 h after the cloud line had been seen near Julia Creek.

The principal conclusions are that long waves of elevation or internal bores propagate over significant distances, that they decay asymptotically into solitary waves and that waves of this type constitute an ubiquitous component in the description of the dynamical state of the lower troposphere.

Finally, we note that solitary wave disturbances may present a severe hazard to aircraft during landing and take-off. Clear-air disturbances of this type normally occur without warning. The vector change in vertical velocity may exceed 16 m s^{-1} in the larger-amplitude events observed at Tennant Creek and Burketown. The hazard will be compounded in the case of an extensive solitary wave family due to the sequence of up- and down-draughts in the flow pattern.

We thank J. S. Turner and R. K. Smith for discussions, J. Tarca, A. Mainwaring and B. Schaefer for operation of the Warramunga Infrasonic Array, and Mount Isa Mines Ltd for data from their environmental meteorological network.

Received 27 February; accepted 30 June 1981.

1. Miles, J. W. *Ann. Rev. Fluid Mech.* **12**, 11–43 (1980); *Tellus* **31**, 456–462 (1979).
2. Hunkins, K. & Fliegel, M. J. *J. geophys. Res.* **78**, 539–548 (1973).
3. Farmer, D. M. & Smith, J. D. in *Hydrodynamics of Estuaries and Fjords* (ed. Nihoul, J.) (Elsevier, Amsterdam, 1978); *Deep-Sea Res.* **27A**, 239–254 (1980).
4. Haury, L. R., Briscoe, M. G. & Orr, M. H. *Nature* **278**, 312–317 (1979).
5. Pinkel, R. J. *phys. Oceanogr.* **9**, 675–686 (1979).
6. Christie, D. R., Muirhead, K. J. & Hales, A. L. *J. atmos. Sci.* **15**, 805–825 (1978); *J. geophys. Res.* **84**, 4959–4970 (1979).
7. Clarke, R. H., Smith, R. K. & Reid, D. G. *Mon. Weath. Rev.* (in the press).
8. Tepper, M. J. *Met. J.* **7**, 21–29 (1950).
9. Freeman, J. C. *Compendium of Meteorology*, 421 (Waverley, Baltimore, 1951).
10. Benjamin, T. B. *J. Fluid Mech.* **25**, 241–270 (1966); **29**, 559–592 (1967).
11. Davis, R. & Acrivos, A. *J. Fluid Mech.* **29**, 593–607 (1967).
12. Ablowitz, M. H., Kaup, D. J., Newell, A. C. & Segur, H. *Studies appl. Math.* **53**, 249–315 (1974).
13. Segur, H. *J. Fluid Mech.* **59**, 721–736 (1973).
14. Hammack, J. L. & Segur, H. *J. Fluid Mech.* **65**, 289–314 (1974); **84**, 337–358 (1978).
15. Joseph, R. I. *J. Phys. A* **10**, L225–L227 (1977); *J. Math. Phys.* **18**, 2251–2258 (1977).
16. Kubota, T., Ko, D. R. S. & Dobbs, L. *J. Hydron.* **12**, 157–165 (1978).
17. Maxworthy, T. *J. Fluid Mech.* **96**, 47–64 (1980).
18. Meiss, J. D. & Pereira, N. R. *Phys. Fluids* **21**, 700–702 (1978).
19. Maslowe, S. A. & Redekopp, L. G. *Geophys. Astrophys. Fluid Dynamics*, **13**, 185–196 (1979); *J. Fluid Mech.* **101**, 321–348 (1980).
20. Grimshaw, R. *Proc. R. Soc. A* **368**, 359–375 (1979); *Math. Res. Rep.* **44** (University of Melbourne, 1980); *Studies appl. Math.* (in the press); *Proc. R. Soc.* (in the press).
21. Matsuno, Y. *J. Phys. A* **12**, 619–621 (1979); *Phys. Lett.* **74A**, 233–235 (1979).
22. Vliegthart, A. C. *J. Engng Math.* **5**, 137–155 (1971).
23. Clarke, R. H. *J. appl. Met.* **11**, 304–311 (1972).
24. Neal, A. B., Butterworth, I. J. & Murphy, K. M. *Weather* **32**, 176–183 (1977).
25. Kao, T. W. & Pao, H.-P. *J. Fluid Mech.* **97**, 115–127 (1979).
26. Robin, A. G. *Aust. Met. Mag.* **26**, 125 (1978).
27. Clarke, R. H. & Brook, R. R. *The Koorin Expedition: Atmospheric Boundary Layer Data over Tropical Savannah Land* (Bureau of Meteorology, Australian Government Publishing Service, Canberra, 1979).
28. Stull, R. B. *J. atmos. Sci.* **30**, 1092–1099 (1973).
29. Carson, D. J. *Q. Jl R. met. Soc.* **99**, 450–467 (1973).

Evidence for northward thrusting south-west of the Rhodes Basin

Derk Jongsma

Free University, Institute for Earth Sciences, 1007 MC Amsterdam, The Netherlands

Jean Mascle

Laboratoire de Géodynamique Sous-Marine, 06230 Villefranche sur-Mer, France

Models of convergence between the major African and European plates involve a stretching Aegean plate, whose present motion relative to Africa is south-west^{1–3}. This motion has resulted in subduction of eastern Mediterranean sea floor as shown by seismicity, volcanism and morphology. The southern boundary of the Aegean plate is formed by the Hellenic Trench. Intensive study, culminating in the use of submersibles and multibeam echo-sounding as part of the Hellenic Arc and Trench Program⁴, has outlined the complexity of this boundary. Detailed reflection profiling which we carried out in autumn 1980 of the eastern portion of the Hellenic Trench system shows arcward verging thrusting south-west of the Rhodes Basin. We conclude that collision of the African continental margin with the Hellenic Arc is now in progress, causing overriding by the Mediterranean Ridge along the southeastern branch of the Hellenic Trench system.

Kinematic analysis³ predicts a compressive branch in the Ionian Sea and a transform branch in the Levantine Sea, south-east of Crete (Fig. 1). Thrust focal mechanisms of earthquakes² and landward dipping reflectors in the inner slope of the Hellenic Trench⁴ are evidence for compression along parts of the margin, while normal faulting and horst and graben structures are prevalent in the areas of predicted transform motion^{6,7}. South of the Hellenic Trench lies the eastern Mediterranean Ridge with a crustal structure similar to an Atlantic-type continental margin^{8–10}. The Rhodes Basin lies to the east of the Hellenic Trench and separates two areas of positive relief: one situated south of Karpathos; the other, the Anaximander Mountains, south of western Turkey. The extension of the Hellenic Trench eastwards as well as the nature of the Anaximander Mountains and surrounding area are poorly understood.

One of the deepest areas in the eastern Mediterranean, the Rhodes Basin (Fig. 1), lies between the seismically active Hellenic Trench system and the almost inactive collision zone of the Anaximander Mountains and Florence Rise^{1,10,11}. To our knowledge refraction data from the Rhodes Basin have not been published and it is therefore not known whether the crust underlying it is oceanic or continental. Its great depth of up to 4.3 km is similar to that of the Ionian Basin, where refraction data⁸ indicate minimum depths to the Moho of 18 km. There is a large negative free air anomaly of up to -240 mGal over the central part of the Rhodes Basin¹⁰. The long-wavelength component ($> 300 \text{ km}$) of this anomaly (see ref. 12, Fig. 8) retains its negative character over the Rhodes Basin where there is still -120 mGal of anomaly suggesting a regionally deep-seated cause. This has suggested¹⁰ that the isostatic imbalance is the result of downward convective flow in the mantle. Previous seismic profiling¹³ shows Pliocene–Quaternary sediments locally thicker than 1 km, overlying a highly disturbed 'M' reflector.

The southern boundary of the Rhodes Basin with the Mediterranean Ridge is formed by an arcuate northward convex structure up to 3.7 km deep. Our profiles shot across the south-western lobe of this area using a 3-l Bolt airgun and a six-element single-channel receiver show $\sim 1 \text{ km}$ of Pliocene–Quaternary sediments in a trough flanking the Mediterranean

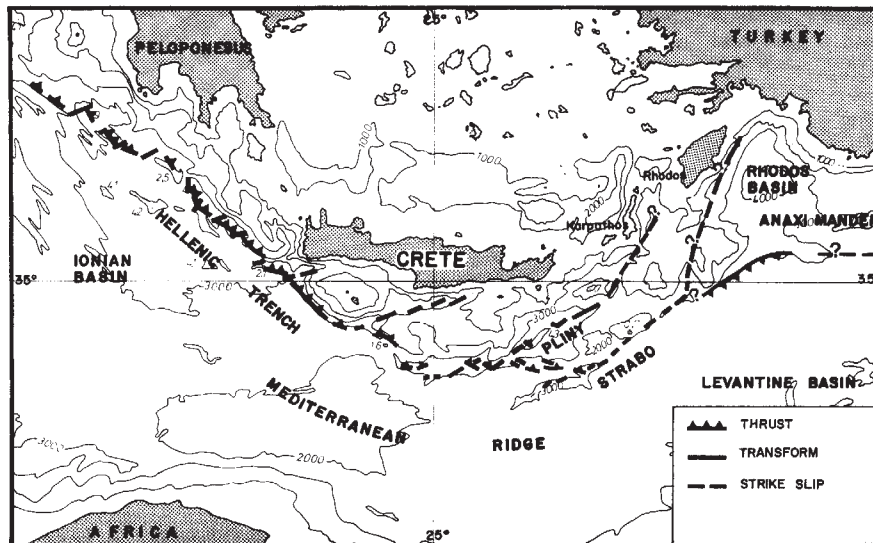


Fig. 1 Tectonic configuration of the Hellenic Trench system as deduced from marine geophysical surveys.

Ridge. This implies a depth of >4.5 km to the base of this trough, almost equalling the deepest area (nearly 5 km) in the Ionian branch of the Hellenic Trench. Further north (10 km) this Pliocene–Quaternary section is only 200–250 m thick. Underlying these sediments is the prominent M reflector associated with the Messinian salinity crisis¹⁴. A second prominent disturbed reflector occurs 1 km deeper in the section. On many of the seismic profiles crossing the trough, the two reflectors continue under the edge of the Mediterranean Ridge and in the deeper part of the Pliocene–Quaternary sequence they tilt slightly northwards.

The seismic reflection profile presented in Fig. 2 convincingly shows that the Pliocene–Quaternary and underlying M reflector continue to some 6 km under the Mediterranean Ridge. Both reflectors are progressively deformed towards the south. We interpret the deformation zone in the north flank of the Ridge as northward thrusting and discount the possibility that the deformation is entirely due to slumping because of the regular upthrusting pattern involving the whole section seen in the profile, from the sea floor to a depth of >1 km.

Almost all known trenches where subduction is taking place exhibit normal faulting in their outer slopes¹⁵. The presence of arcward thrusts in the outer slope of the Hellenic Trench as shown by our seismic profiles and reported¹⁶ further west in the outer slope of the Pliny Trench from observations made with a submersible shows that additional factors to those associated with normal subduction or transform movement must be involved. We propose that this arcward thrusting is due to overriding of the Mediterranean Ridge sediments over the forearc in a manner analogous to that suggested for the Talaud Ridge in the Molucca Sea¹⁷. The Talaud Ridge lies between the two converging active Halmahera and Sangihe volcanic arcs, and is on both sides being thrust onto the flanking arc aprons along gently dipping reverse faults which crop out in troughs adjacent to the arcs. Our explanation is supported by the similarity in the geometry of the reflectors in our profiles and those in the reflection seismic profiles run across the flanks of the Molucca Sea collision zone¹⁷. In both areas reflectors below the inner slope are seen to continue below the outer slope of the troughs^{6,17}. (See ref. 18, Fig. 5 for a good example of reflectors dipping southward below the northern flank of the Mediterranean Ridge south of the western tip of Crete.)

The plate boundary shown in Fig. 1 is derived from a compilation of available seismic profiling data^{5–7} across the Hellenic Trench. In our survey area, a non-sedimented inner slope associated with the Strabo Trench clearly contrasts a thickly sedimented slope south of the marginal relief off Karpachos. (See also ref. 6, Fig. 4.) Local tectonic activity is deduced

from northward dipping reflectors within the Strabo Trench inner slope, tectonic mixing encountered during deep-sea drilling (DSDP Site 129¹⁴) and direct observations from a submersible¹⁶. We conclude that the active boundary represented by the Strabo Trench terminates in a triangular-shaped embayment within the marginal relief off Karpachos (Figs 1 and 3). Strong seismicity below the marginal relief off Karpachos (Fig. 3) and shallow NNE dipping thrusts in this area² may indicate continuation of this structural boundary east of Rhodes (Fig. 1).

South of Rhodes (Fig. 3) a complicated junction exists between the transform part of the Strabo Trench and the southwestern slope of the Rhodes Basin. The eastern extension of the proposed northward thrusting zone requires further study to establish its structural relation with the Anaximander Mountains. A published profile¹² shows a small cleft south of a steep scarp with a northward-tilted sedimentary basin. Southward thrusting is postulated¹² there.

Southward thrusting is also documented² for the northwestern margin of the Rhodes Basin by a focal mechanism solution of an earthquake which occurred below the Turkish coast in 1969.

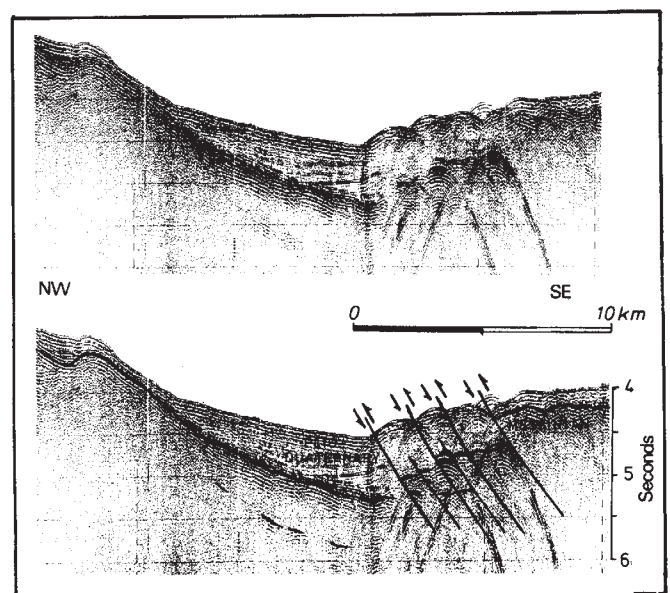


Fig. 2 Seismic reflection profile across the southwestern margin of the Rhodes Basin. Location of the profile is shown in Fig. 3.

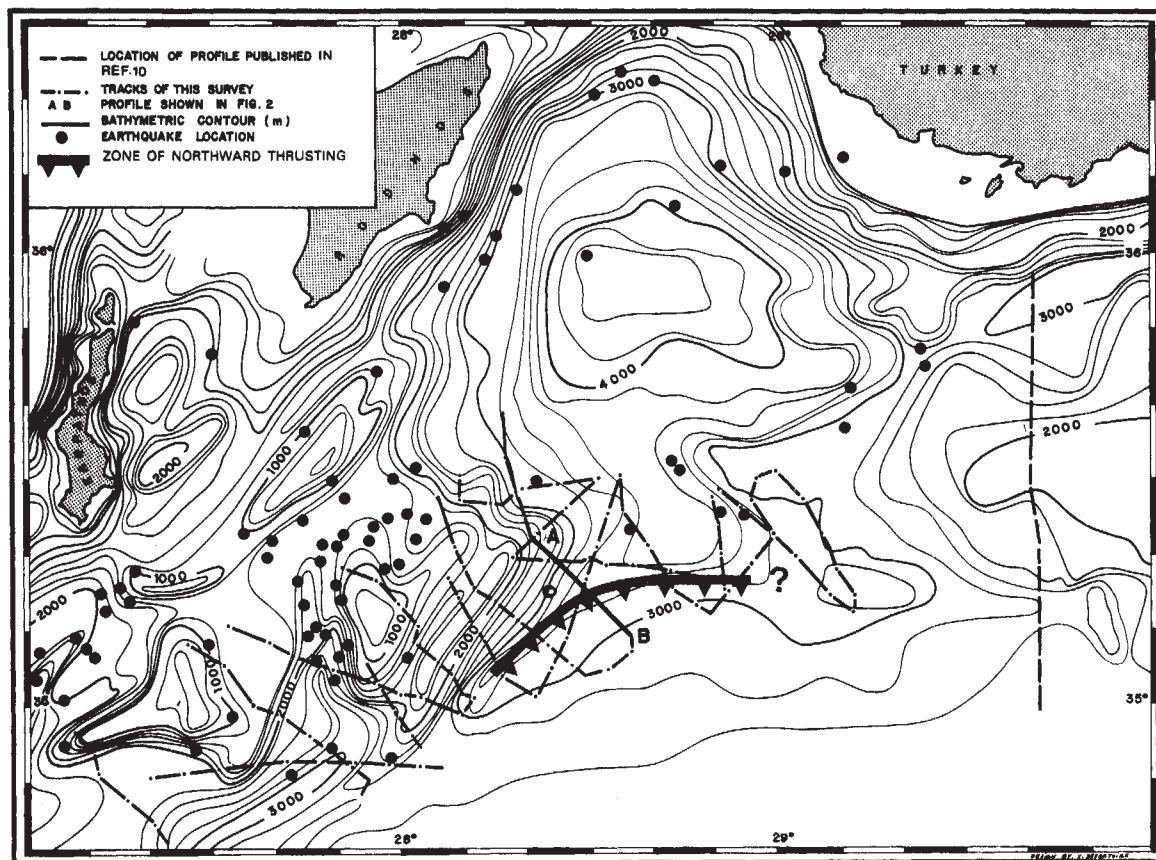


Fig. 3 Detailed bathymetry²¹ and seismicity² of the area surrounding the Rhodes Basin. Reflection seismic profiling tracks and location of zone of northward thrusting as indicated.

This combination of opposite-directed thrusting on either side of the Rhodes Basin may contribute to the isostatic imbalance of the basin reflected in the large negative free air gravity anomaly.

Our interpretation of arcward thrusting due to overriding of the Mediterranean Ridge sediments onto the Hellenic forearc solves several problems about the nature of subduction in the Hellenic Trench and the nature of the Ridge. It explains why the usual configuration of reflectors dipping from below the trench under the inner slope¹⁵ has not been observed in the numerous seismic reflection profiles of this boundary. It would also explain the proposed presence of Messinian evaporites as diapirs within the trench axis¹⁹. Finally, the observed folding¹² and interpreted southward thrusting in seismic profiles²⁰, crossing the southern flank of the Mediterranean Ridge, are compatible with our interpretation of this structure as a collision complex similar to the Talaud Ridge. The Talaud Ridge is, however, at a more advanced stage of collision than the Mediterranean Ridge.

We acknowledge the assistance of colleagues and crew of the *N/O Le Noroît*, and CNRS and CRM for financial support.

Received 8 April; accepted 8 June 1981.

1. McKenzie, D. P. *Geophys. J.R. astr. Soc.* **30**, 109–185 (1972).
2. McKenzie, D. P. *Geophys. J.R. astr. Soc.* **58**, 217–254 (1978).
3. Le Pichon, X. & Angelier, J. *Tectonophysics* **60**, 1–42 (1979).
4. Le Pichon, X. & Hsü, K. J. *Episodes* **1**, 3–6 (1979).
5. Le Quellec, P. et al. *Bull. Am. Ass. petrol. Geol.* **64**, 242–263 (1980).
6. Jongsma, D. *Bull. geol. Soc. Am.* **88**, 797–805 (1977).
7. Leite, O. et al. *C.r. somm. Séanc. Soc. géol. Fr.* **5**, 171–174 (1980).
8. Makris, J. *Tectonophysics* **46**, 251–289 (1978).
9. Cloething, S. et al. *Earth planet. Sci. Lett.* **51**, 336–342 (1980).
10. Woodside, J. M. *Mar. geophys. Res.* **3**, 317–354 (1977).
11. Bijou-Duval, B. et al. *Init. Rep. DSDP Leg 42*, 951–984 (1978).
12. Woodside, J. M. *Geophys. J. R. astr. Soc.* **47**, 493–514 (1976).
13. Wong, H. K. *Bull. geol. Soc. Am.* **82**, 91–100 (1971).
14. Hsü, K. S. et al. *Nature* **242**, 240–244 (1973).
15. Dickinson, W. R. & Seely, D. R. *Bull. Am. Ass. petrol. Geol.* **63**, 2–31 (1979).
16. Le Pichon, X. et al. *C.r. heb. Séanc. Acad. Sci., Paris* **289** (D), 1225–1228 (1979).
17. Silver, E. A. & Moore, J. C. *J. geophys. Res.* **83**, 1681–1691 (1978).
18. Sancho, J. et al. *Earth planet. Sci. Lett.* **18**, 189–204 (1973).
19. Le Pichon, X. et al. *C.r. heb. Séanc. Acad. Sci., Paris* **290** (D), 5–8 (1980).
20. Finetti, I. *Bull. Geol. teor. appl.* **13**, 31–65 (1976).
21. Wright, D. W. et al. *Boll. Geofis. teor. appl.* **18/66**, 168 (1975).

Accelerated currents and sediment transport off the Damietta Nile promontory

Stephen P. Murray, James M. Coleman
& Harry H. Roberts

Coastal Studies Institute, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Monir Salama

Coastal Protection Branch, Institute of Oceanography and Fisheries, Alexandria, Egypt

Severe erosion of valuable agricultural and recreational lands along the apex of the Damietta Nile promontory (Fig. 1) following the closure of the Aswan High Dam remains a problem. Despite extensive shoreline studies, the fate of eroded material and the main mechanism of sediment transport along this coast are unclear¹. An 11-day bathymetric survey provided exploratory measurements in the coastal waters surrounding the promontory. We report here current meter and side-scan sonar observations which indicate that a high-speed jet forming off the promontory is driving a field of actively migrating sand ridges easterly over a smooth mud plain. These ridges seem to be the dominant along-shore pathway for sediment transport between Damietta and the Suez Canal.

Tidal data, including our own measurements, indicate a 10-cm average range and a 5-cm s⁻¹ current amplitude. Wind direction is northerly from May to October and southwesterly from December to March, with speeds of 4–10 m s⁻¹. The background current sets easterly, parallel to the coast throughout the year at 10–15 cm s⁻¹, except for brief flow reversals in spring and late autumn².

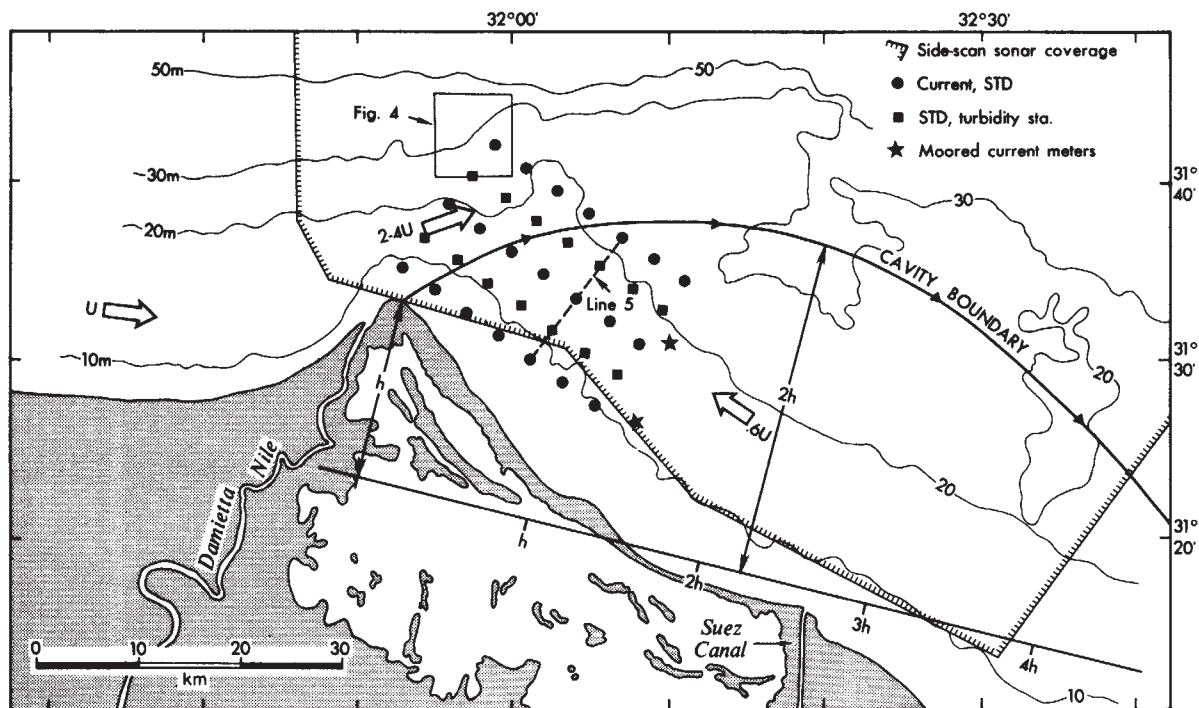


Fig. 1 Location map of Damietta promontory showing data coverage in 1978. Side-scan sonar coverage extends over the entire study area out to the shelf edge.

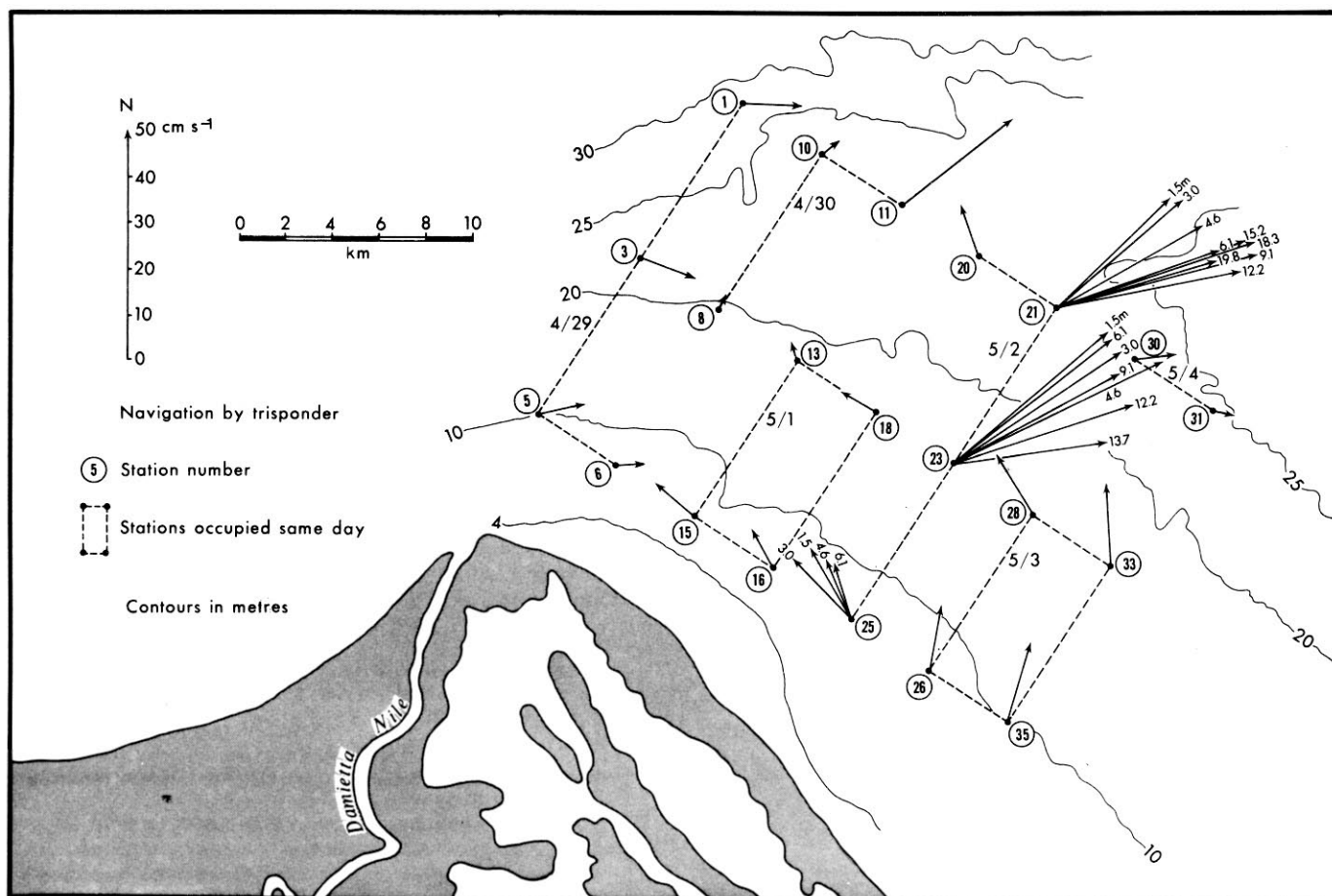


Fig. 2 Vertically averaged current observations except at stations 21, 23 and 25, which show the complete vertical current profiles. On 1 and 2 May winds were from the north-east at $5\text{--}10\text{ m s}^{-1}$.

Early charts of the Damietta promontory (for example, USHO 3976) included undated Admiralty surface current vectors showing an eddy in the embayment east of the promontory. Strong ($50\text{--}90\text{ cm s}^{-1}$) north-east-setting currents off the apex turn southeasterly to flow (50 cm s^{-1}) parallel to the coast, 30 km offshore. Vectors (35 cm s^{-1}) turn directly towards the coast 35 km east of Damietta, and a lower speed return flow ($12\text{--}20\text{ cm s}^{-1}$) forms the inshore limb of an elongate eddy.

Using a 12-m launch we sequentially occupied a 35-station grid (Fig. 1) in a 6-day period. Vertical profiles of current velocities were taken on alternate lines, while STD casts (Plessey 9060) for density were taken at every station. Two bi-level current meter moorings were deployed in the onshore flow zone indicated by the Admiralty data.

Figure 2 shows the vertically averaged current velocities at all grid stations except 21, 23 and 25, where the complete profile illustrates the depth and intensity of an outflow jet east of the promontory. Transient wind effects near the surface³ are largely eliminated by deleting the upper two levels (1.5 and 3.0 m) from the integration.

Note that stations 20–25 were taken in sequence at 1-h intervals. The strong shear between 20 and 21 and between 23 and 25 suggests a relatively narrow, intense jet observed to extend throughout the water column. The jet moved seaward against the local wind, which was from north to north-east at $5\text{--}10\text{ m s}^{-1}$ from 30 April to 2 May. Despite distortion due to lack of synopticity and some meandering of the outflow zone, the observations, namely, (1) the return flow along the coast; (2) a stagnation zone; (3) a flow separation north-east of the apex, and (4) the outflow jet, delineate the western edge of a clockwise eddy about twice the size of that described by the Admiralty data. The moored current meters, due to their interior position, failed to measure the main current stream. However, the records do show the tidal current amplitude to be only $\sim 5\text{ cm s}^{-1}$, indicating that the serial anchor stations observed a reasonably steady current.

Although the high-speed jet ran seaward beyond the grid, the side-scan sonar provided evidence of its track. A previously

undiscovered belt of sand ridges (0.3-mm median diameter) actively migrating over a smooth muddy bottom was found to trend seaward (Fig. 3) off the Damietta mouth. These long, linear sand ridges at depths of 25–60 m rise 2–8 m above the sea floor and on top of these are mobile sand waves (Fig. 4). Sand-wave movement (indicated by slip face orientation, Fig. 3) is northeasterly near the Damietta mouth, turns easterly and finally arcs southeasterly towards the coast off the Suez Canal.

Migrating sand waves demand speeds much higher than the $\sim 30\text{ cm s}^{-1}$ necessary to initiate motion of the 0.3-mm sand ridges⁴. Field and laboratory data on channel flow^{5,6} show that a current stream power, $\tau_0 \bar{u}$ (where $\bar{u}$ is channel average velocity and τ_0 is boundary shear stress) of $\sim 1,500\text{ erg s}^{-1}\text{ cm}^{-2}$ is required to generate sand-wave fields. For the shelf⁷ a scaling value of $u_*/\bar{u} \approx 1/30$ is obtained, where $\bar{u}$ is a representative average velocity and u_* is friction velocity. Together with $\tau_0 = \rho u_*^2$, these relationships indicate that u_* values of $\sim 4\text{ cm s}^{-1}$ and $\bar{u}$ values of $\sim 100\text{ cm s}^{-1}$ are necessary to drive sand waves. These estimates suggest that the sand-ridge complex is driven by currents even stronger than those observed off the promontory (Fig. 2). Extension of the observed current jet out over the sand-ridge complex seems likely, but more detailed observations are needed to confirm this.

The exploratory nature of the observations allows only a scale analysis of the momentum balance. Values of 50 cm s^{-1} for a current speed scale U , 50 m for a vertical length scale H , and 50 km for a horizontal length scale L are appropriate. Scale analysis of the momentum equation yields the baroclinic pressure gradient term, scaled as $(g/\rho)(\Delta\rho/L)$, $H = 6 \times 10^{-4}\text{ cm s}^{-2}$, the inertial term $U^2/L = 5 \times 10^{-4}\text{ cm s}^{-2}$, and the Coriolis term $fU = 4 \times 10^{-3}\text{ cm s}^{-2}$. Fluctuations in the moored current meter data suggest $\Delta u/\Delta t$ to be $0(10^{-4})\text{ cm s}^{-2}$. If the friction term is of the order of the nonlinear inertial term⁸, then the barotropic pressure gradient term gS (S is sea-surface slope) provides the major balance to the Coriolis term. The large-scale flow thus seems to be quasi-geostrophic. Locally, for example, near the promontory, nonlinear effects dominate and there may be a considerable slope in sea level. For example, Pingree⁹ shows a

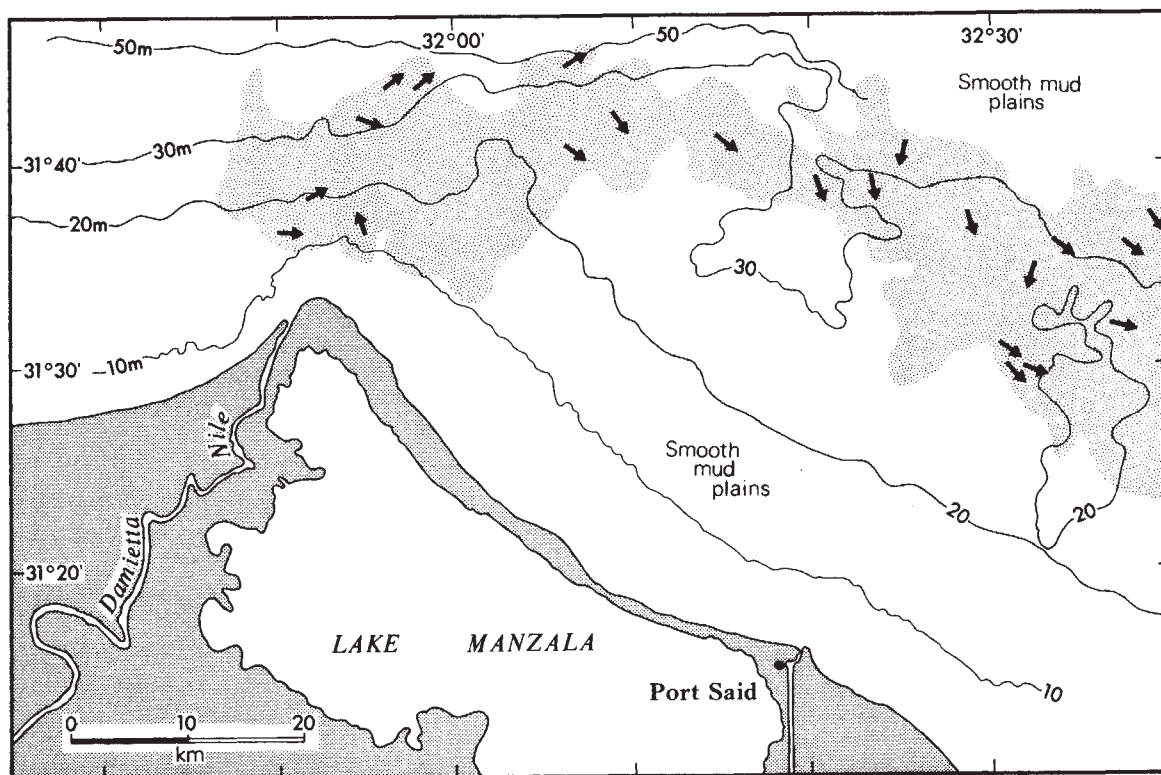


Fig. 3 The sand-ridge field off the Damietta promontory mapped by side-scan sonar, 150-m line separation, 50-m overlap from the 10-m contour out to the shelf edge. Direction of movement of sand waves is indicated by arrows. (See detail in Fig. 4.)

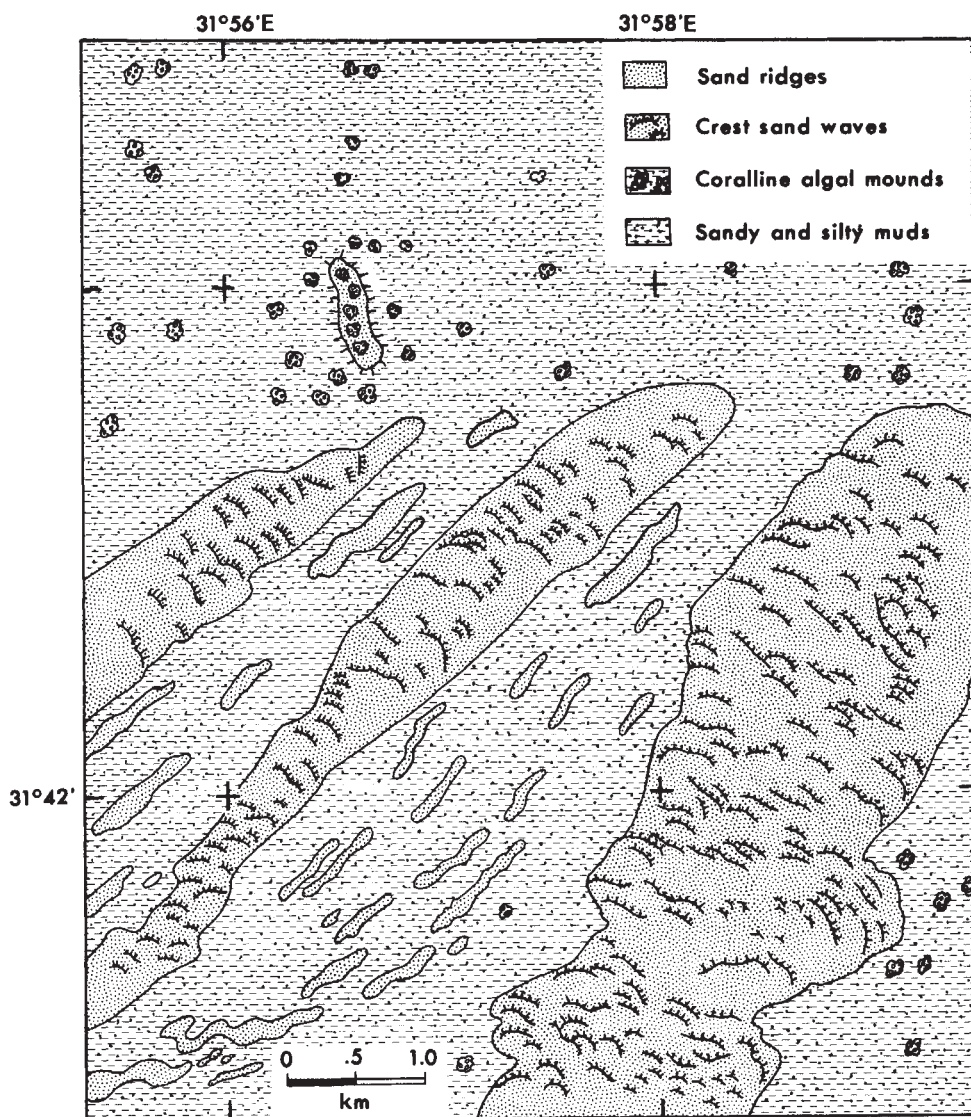


Fig. 4 Detailed side-scan sonar map of edge of sand-ridge field showing individual sand ridges and superimposed migratory bedforms. For location see Fig. 1.

tidal residual barotropic pressure gradient of almost $10^{-1} \text{ cm}^2 \text{ s}^{-1}$ confined to within 10 km of the Portland headland.

The control of residual eddies on sediment accumulation forms has recently been investigated numerically⁹ in the field¹⁰ and in the laboratory¹¹. Ferentino and Collins¹⁰ argued that sediment will accumulate passively in the stagnant central region of eddies. In contrast, Pingree⁹ demonstrated that all eddies do not have the bottom boundary layer (BBL) convergence that leads to sediment accumulation at their centres. Eddies of small Rossby number (ratio of inertial forces to Coriolis forces) $R_0 \ll 1$ and clockwise rotation will have a central high-pressure region and a divergent BBL not conducive to central sediment accumulation. R_0 of only 0.1 and clockwise rotation east of the Damietta may thus explain the lack of a centralized sediment deposit. The possibility that the Damietta ridges initially accumulate in a convergent BBL expected⁹ along the outer periphery of a clockwise eddy of low R_0 should be evaluated.

Note also that the curvature and size of the sand-ridge field and the direction of sand-wave movement (Fig. 3) generally resemble the expected shapes of eddies that develop behind obstacles to unidirectional flow¹²⁻¹⁴. Commonly observed features shown in Fig. 1 are an eddy length five to six times the obstacle size, a two- to four-fold amplification of upstream speed around the obstacle, and a broad, slow return flow of magnitude 0.5–0.6 of the upstream flow.

Thus, our exploratory measurements indicate that a newly discovered sand-ridge system driven by strong topographically induced currents seems to be a major sink for the products of the

extensive coastal erosion occurring along the Damietta promontory. Future work should include trajectory studies and current meters moored off the promontory to determine temporal and spatial behaviour of the jet and its relation to both the sand-ridge field and a large eddy ($L \approx 50 \text{ km}$) that apparently occurs east of the promontory. The fate of the sand at the east end of the sand-ridge system is another point of interest.

This work was supported by the Coastal Sciences Program of the US Office of Naval Research and the US Naval Oceanographic Office. The professional assistance of the crew of the survey boats attached to the USNS *Harkness* is gratefully acknowledged.

Received 1 June; accepted 15 July 1981.

1. United Nations Development Program Proc. Semin. Nile Delta Coastal Processes, 624 (Acad. Sci. Res. Tech., Cairo, 1977).
2. Gerges, M. IV Journées Etud. Pollutions, Antalya, 573–587 (CIESM, 1978).
3. Murray, S. P. *J. phys. Oceanogr.* **5**, 347–360 (1975).
4. Yalin, M. S. *Mechanics of Sediment Transport*, 290 (Pergamon, New York, 1972).
5. Graf, W. H. *Hydraulics of Sediment Transport*, 513 (McGraw-Hill, New York, 1971).
6. Simons, D. B. & Richardson, E. V. *Resistance to Flow in Alluvial Channels*, Pap. 422-J, 61 (US Geol. Survey, Washington, 1966).
7. Wimbush, M. in *The Benthic Boundary Layer* (ed. McCave, I. N.) 323 (Plenum, New York, 1976).
8. Pond, S. & Pickard, G. L. *Introductory Dynamic Oceanography*, 241 (Pergamon, New York, 1978).
9. Pingree, R. D. *J. mar. Biol. Ass. U.K.* **58**, 211–226 (1978).
10. Ferentino, M. & Collins, M. *J. sedim. Petrol.* **50**, 1081–1094 (1980).
11. Sugimoto, T. *J. oceanogr. Soc. Japan* **31**, 1–14 (1975).
12. Halitsky, J. in *Meteorology and Atomic Energy* (ed. Slade, R. H.) 221–255 (US Atomic Energy Commn, Tennessee, 1968).
13. Plate, E. J. *Aerodynamic Characteristics of Atmospheric Boundary Layers*, 190 (US Atomic Energy Commn, Tennessee, 1971).
14. Mason, P. J. & Sykes, R. I. *Q. J. R. met. Soc.* **105**, 129–146 (1979).

Testis weight, body weight and breeding system in primates

A. H. Harcourt*, P. H. Harvey†, S. G. Larson‡ & R. V. Short§

*Department of Applied Biology, University of Cambridge, Pembroke St, Cambridge CB2 3DX, UK

†School of Biological Sciences, University of Sussex, Brighton, Sussex BN1 9QG, UK

‡Department of Anthropology, University of Wisconsin, Madison, Wisconsin 53706, USA

§MRC Reproductive Biology Unit, 37 Chalmers St, Edinburgh EH3 9EW, UK

It has long been known that primate species differ greatly in the weight of their testes relative to body weight¹. Recently it has been suggested that among the three species of Pongidae (the great apes), the disparity in testes weights is associated with their different breeding systems²⁻⁴. Male gorillas and orangutans copulate infrequently, and when a female comes into oestrus she normally mates with only one male. However, in the chimpanzee, several males mate frequently with the oestrous females, so that each male has to deposit enough sperm to compete with the presence of sperm from other males. For the chimpanzee, therefore, we hypothesize that selection will favour the male that can deposit the largest number of sperm; thus the volume of spermatogenic tissue and hence testis size is far greater in the chimpanzee than in the gorilla or orangutan. If this is correct, it implies that primates in which more than one male mates with each oestrous female should have larger testes relative to their body weight than those with single-male breeding systems. We have tested this prediction across a wide range of primates, and the results support the hypothesis. The relative size of testes may, therefore, provide a valuable clue to the breeding system of a primate species.

The male gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*), heaviest of the primates, have breeding systems that involve one male monopolizing mating with a number of females⁵⁻⁷, and have testes that together weigh 30 g and 35 g respectively. The lighter chimpanzee (*Pan troglodytes*) male, by contrast, has a breeding system in which ~75% of copulations and at least 25% of conceptions occur during periods of promiscuous mating when several males copulate with each oestrous female⁸ (multi-male breeding system); its combined testis weight is ~120 g. It has been suggested that the marked differences in testes weights among the three species are related to their breeding systems²⁻⁴. In the single-male breeding system of the gorilla and orangutan, each male need ejaculate only enough sperm to ensure fertilization, whereas in the multi-male system of the chimpanzee, each male also has to inseminate sufficient to compete with the presence of sperm from other males. Thus, selection in the chimpanzee for high sperm production may explain why it has larger testes than the gorilla or orangutan. High sperm storage could also be selected for, but we have too few data to test this idea.

An obvious test of the hypothesis relating size of testes to breeding system is to determine whether a similar association is found in other primates. This has been done for baboons⁹, but the sample size was small and allometric effects were not taken into account. We therefore examined a wide range of primate species representing 18 genera from six families, from a 320-g marmoset (*Callithrix*) to the 170-kg gorilla. Data from 33 species taken from the literature and some unpublished results are shown in Table 1. No prosimian species were included in the statistical analysis because of lack of reliable information on their mating systems. Only data from fully mature individuals were included, and the body weights are those of the individuals whose testes were weighed. Where more than one specimen per species was available, the testes weight of the individual of median body weight was used, except for *Saimiri sciureus* and

Presbytis entellus where the original sources required use of, respectively, the mean and the midpoint between extreme values. Note that throughout we considered breeding, not social, systems. The distinction is relevant for species such as the gorilla⁵, that can occur in multi-male groups but have single-male breeding systems.

Testes weight increases with body weight, even for species with the same breeding system (Fig. 1). We therefore corrected for the effect of body weight by examining deviations from the line of best fit. The major axis line was used because some error was incorporated into measures of both body weight and testes weight, and because the data were logarithmically transformed¹⁰. The taxonomic level analysed was the genus. Generic values were calculated from the means of the logarithmically transformed values for species, but congenics were kept separate if they had different breeding systems. The species level was not used in statistical analysis because closely related species within various genera have very similar values and cannot be treated as independent points. In addition, when the species level was analysed, we found significant heterogeneity in deviations due to family membership (taxonomic effect) within the monogamous group. Thus, analysis at the genus level allowed us to examine the relationship between testes weight and breeding system, with body weight and taxonomic effects removed.

As no significant heterogeneity in slope among breeding systems was revealed by maximum likelihood analysis¹⁰ ($\chi^2_2 = 3.10$), a common slope of 0.66 was fitted through all the points (Fig. 1). Furthermore, within breeding systems, there was no heterogeneity of deviations from the line among taxa (monogamous, $F_{2,1} = 2.57$; single male, $F_{2,4} = 2.43$; multi male, $F_{2,5} = 3.33$; one-way analysis of variance). Among breeding systems, however, as predicted, we found significant heterogeneity of deviations from the common slope ($F_{2,16} = 10.29$; $P < 0.01$); whereas monogamous and single-male genera do not differ from each other in their deviations about the common line ($t_5 = 0.62$; Student's *t*-test), both show the predicted significant differences from the multi-male genera ($t_6 = 2.45$, $P < 0.05$ and $t_{15} = 4.76$, $P < 0.01$, respectively). As the deviations were measured perpendicular to the body weight axis, the results show that multi-male genera have significantly heavier testes in relation to body weights than do the single-male or monogamous genera.

The positive correlation between testes weight and body weight, having a common slope of < 1.0 , is true of many organ to body weight relationships among mammals¹¹. Also, in this study a significant correlation between the two variables was found in three of the eight species for which records from > 10 specimens were available (median $N = 17.5$; $P < 0.05$, Spearman rank

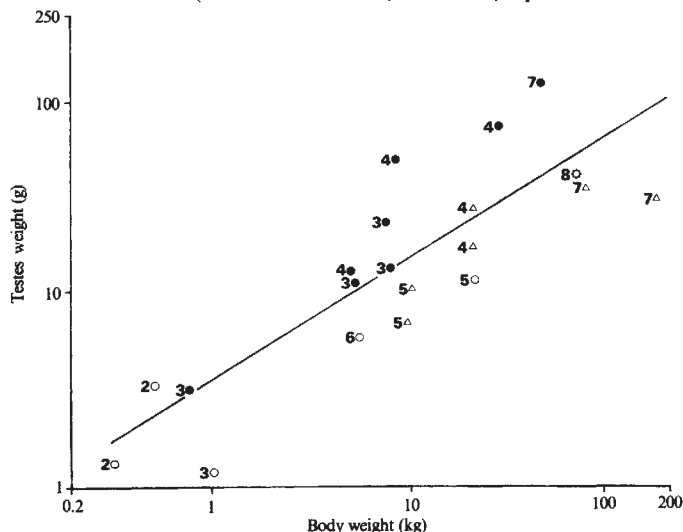


Fig. 1 Log combined testes weight (g) versus log body weight (kg) for different primate genera. ●, Multi-male breeding system; ○, monogamous; △, single-male; ○, *Homo*. Numbers represent the families shown in Table 1. (4) and (5) were combined for statistical analysis.

Table 1 Combined testes weight (excluding epididymis), body weight and mating system of primates

	N	Body weight (kg)	Testes weight (g)	Ratio (%)	Mating system	Refs Weights	Mating system
(1) Lorissidae							
<i>Loris tardigradus</i> (w)	114	0.27	1.8	0.66	?	14	
(2) Callitrichidae							
<i>Callithrix jacchus</i> (c)	15	0.32	1.3	0.41	P	*	15
<i>Sanguinus oedipus</i> (w)	56	0.52	3.4	0.65	P	1, 16	17
(3) Cebidae							
<i>Saimiri sciureus</i> (w)	40	0.78	3.2	0.41	M	18	19
<i>Aotus trivirgatus</i> (w)	1	1.02	1.2	0.12	P	1	15
<i>Lagothrix lagothricha</i> (c)	1	5.22	11.2	0.2	M	1	15, 20
<i>Alouatta palliata</i> (w)	8	7.26	23.0	0.32	M	21	22
<i>Ateles geoffroyi</i> (w)	1	7.94	13.4	0.17	M	1	15
(4) Cercopithecinae							
<i>Cercopithecus aethiops</i> (w)	6	4.95	13.0	0.26	M	23	24
<i>Cercocebus atys</i> (c)	1	8.68	25.1	0.29	?	25	
<i>Macaca fascicularis</i> (w, c)	20	4.42	35.2	0.80	M	1, †	26
<i>Macaca radiata</i> (w, c)	2	8.65	48.2	0.56	M	1, ‡	27
<i>Macaca mulatta</i> (c)	26	9.2	46.2	0.50	M	1, 12, ‡	28
<i>Macaca nemestrina</i> (w)	1	9.98	66.7	0.67	M	1	29
<i>Macaca arctoides</i> (c)	20	10.51	48.15	0.46	M	‡	29
<i>Papio hamadryas</i> (w)	6	20.17	27.1	0.13	S	23	30
<i>Papio cynocephalus</i> (c)	21	24.32	52.0	0.21	M	‡	31
<i>Papio anubis</i> (w)	4	26.40	93.5	0.35	M	23	32
<i>Papio ursinus</i> (?)	1	31.75	72.0	0.23	M	33	34
<i>Papio papio</i> (w)	1	31.98	88.9	0.28	M	1	35
<i>Theropithecus gelada</i> (w)	1	20.40	17.1	0.08	S	23	36
(5) Colobinae							
<i>Presbytis rubicunda</i> (w)	12	6.23	3.4	0.05	?	1	
<i>Presbytis cristata</i> (w)	12	6.58	6.2	0.09	S	1	37
<i>Presbytis obscura</i> (w)	14	7.45	4.8	0.06	S	§	38
<i>Presbytis entellus</i> (w)	6	17.0	11.1	0.06	S	39, 40	41
<i>Colobus polykomos</i> (= <i>guereza</i>) (w)	3	10.25	10.7	0.10	S	19	42
<i>Nasalis larvatus</i> (w)	8	20.64	11.8	0.06	S	1	43,
(6) Hylobatidae							
<i>Hylobatus moloch</i> (w)	7	5.44	6.1	0.11	P	1	44
<i>Hylobates lar</i> (c)	1	5.5	5.5	0.10	P	45	44
(7) Pongidae							
<i>Pan troglodytes</i> (c)	3	44.34	118.8	0.27	M	1	8
<i>Pongo pygmaeus</i> (w)	2	74.64	35.3	0.05	S	1	6, 7
<i>Gorilla gorilla</i> (w)	2	169.0	29.6	0.02	S	46, 47	5
(8) Hominidae							
<i>Homo sapiens</i>	4	65.65	40.5	0.06	S, P	1, 48	49

(w), Specimens caught from the wild; (c), captive specimens. P, monogamous (pair); M, multi-male; S, single-male breeding system. Unpublished data from *R.V.S.; †P. Squires; ‡S.G.L.; §G. J. Burton and Earl Cranbrook; ||M. Cutler and S. M. Jeffrey.

correlation coefficients). Among species, such a correlation is expected because the testes are endocrine glands whose output and hence volume has to increase with body size if threshold concentrations of hormones are to be maintained. In addition, larger species require more sperm to counteract the dilution effect of the larger volume of the male and female reproductive tracts⁴.

The hypothesis that competition in sperm numbers is important in the association of testes size with breeding system requires that the multi-male primates, as well as having larger testes, have a greater volume of seminiferous tubules, rather than interstitial tissue, and greater sperm production capabilities than the one-male or monogamous species. This seems to be the case; *Pan*, *Papio* (baboon) and *Macaca* (macaque), which all have multi-male breeding systems, have ratios of tubules to connective tissue ranging from 2.2 to 2.8:1, whereas *Homo*, *Hylobates* (gibbon) and *Presbytis* (langur), which are all non-multi-male species, have ratios of only 0.9 to 1.3:1 (ref. 1). In addition, the sperm production rate of the multi-male rhesus macaque (*Macaca mulatta*), with its high volume of tubules, is 23×10^6 sperm per g of testis per day, whereas in man it is only 4.4×10^6 sperm per g per day^{12,13}.

Differences in breeding system are clearly not the only reason

for variation in testes weight among the primates. Seasonality of breeding, for example, could be an important factor: species with a short breeding season, which have a high level of copulatory activity for a period, probably need larger testes. Certainly the multi-male genus *Macaca*, a predominantly seasonal breeder, shows the largest deviation from the common line. Nevertheless, there is clearly a relationship between testes weight and breeding system, despite exceptions such as *Saguinus oedipus* (tamarin); thus it seems reasonable to predict that similar trends will be found in other mammalian orders.

We thank Drs J. Clevedon Brown, T. H. Clutton-Brock, R. D. Martin and B. P. Setchell for valuable advice and criticism; and the following for unpublished data: Dr G. J. Burton, Earl Cranbrook and the University of Malaya, P. Squires, M. Cutler, S. M. Jeffrey and the Huntingdon Research Centre.

Received 6 January; accepted 13 July 1981.

- Schultz, A. H. *Anat. Rec.* **72**, 387-394 (1938).
- Short, R. V. in *Reproduction and Evolution* (eds Calaby, J. H. & Tyndale-Biscoe, C. H.) (Australian Academy of Science, Canberra, 1977).
- Short, R. V. *Adv. Study Behav.* **9**, 131-158 (1979).
- Short, R. V. in *Reproductive Biology of the Great Apes* (ed. Graham, C. E.) (Academic, New York, 1981).
- Harcourt, A. H. in *Reproductive Biology of the Great Apes* (ed. Graham, C. E.) (Academic, New York, 1981).

6. Rijksen, H. D. *A Fieldstudy on Sumatran Orang Utans* (Veenman & Zonen, Wageningen, 1978).
7. Galdikas, B. M. F. in *Reproductive Biology of the Great Apes* (ed. Graham, C. E.) (Academic, New York, 1981).
8. Tutin, C. E. G. *J. Reprod. Fert. Suppl.* **28**, 43–57 (1980).
9. Collins, A. *New Scientist* **78**, 12–14 (1978).
10. Harvey, P. H. & Mace, G. M. in *Current Problems in Sociobiology* (ed. King's College Research Centre Sociobiology Project) (Cambridge University Press, in the press).
11. Brody, S. *Bioenergetics and Growth* (Reinhold, New York, 1945).
12. Amann, R. P., Johnson, L., Thompson, D. L. & Pickett, B. W. *Biol. Reprod.* **15**, 586–592 (1976).
13. Amann, R. P. & Howards, S. S. *J. Urol.* **124**, 211–215 (1980).
14. Ramakrishna, P. A. & Prasad, M. R. N. *Folia primatol.* **5**, 176–189 (1967).
15. Moynihan, M. *The New World Primates* (Princeton University Press, New Jersey, 1976).
16. Dawson, G. A. & Dukelow, W. R. *J. med. Primatol.* **5**, 266–275 (1976).
17. Moynihan, M. *Smithson. Contr. Zool.* **28**, 1–77 (1970).
18. Middleton, C. C. & Rosal, J. *Lab. Anim.* **22**, 583–586 (1972).
19. Baldwin, J. D. *Folia primatol.* **9**, 281–314 (1968).
20. Kavanagh, M. & Dresdale, L. *Primates* **16**, 285–294 (1975).
21. Hrdlicka, A. *Am. J. phys. Anthropol.* **8**, 201–211 (1925).
22. Chivers, D. J. *Folia primatol.* **10**, 48–102 (1969).
23. Kinsky, M. *Anat. Anz.* **108**, 65–82 (1960).
24. Struhsaker, T. T. *Behaviour* **29**, 83–121 (1967).
25. Kennard, M. A. & Willner, M. D. *Endocrinology* **28**, 977–984 (1941).
26. Angst, W. in *Primate Behavior. Developments in Field and Laboratory Research* Vol. 4 (ed. Rosenblum, L. A.) (Academic, New York, 1975).
27. Simonds, P. E. in *Primate Behavior. Field Studies of Monkeys and Apes* (ed. DeVore, I.) (Holt, Rinehart & Winston, New York, 1965).
28. Southwick, C. H., Beg, M. A. & Siddiqi, M. R. in *Primate Behavior. Field Studies of Monkeys and Apes* (ed. DeVore, I.) (Holt, Rinehart & Winston, New York, 1965).
29. Roonwal, M. L. & Mohnot, S. M. *Primates of South Asia* (Harvard University Press, Massachusetts, 1977).
30. Kummer, H. *Social Organisation of Hamadryas Baboons* (University of Chicago Press, 1968).
31. Hausfater, G. *Contr. Primatol.* **7**, 1–150 (1975).
32. Hall, K. R. L. & DeVore, I. in *Primate Behavior. Field Studies of Monkeys and Apes* (ed. DeVore, I.) (Holt, Rinehart & Winston, New York, 1965).
33. Hill, W. C. O. *Primates. Comparative Anatomy and Taxonomy 8, Cynopithecinae* (Edinburgh University Press, 1970).
34. Hall, K. R. L. *Proc. zool. Soc. Lond.* **139**, 283–327 (1962).
35. Dunbar, R. I. M. & Nathan, M. F. *Folia primatol.* **17**, 321–334 (1972).
36. Dunbar, R. I. M. & Dunbar, P. *Contr. Primatol.* **6** (1975).
37. Bernstein, I. S. *Behaviour* **32**, 1–16 (1968).
38. Curtin, S. H. thesis, Univ. Calif., Berkeley (1976).
39. David, G. F. X. & Ramaswami, L. S. *J. Morph.* **135**, 99–130 (1971).
40. Braz, I., Shandilya, L. N. & Ramaswami, L. S. *Andrologia* **8**, 290–296 (1976).
41. Hrdy, S. B. *The Langurs of Abu* (Harvard University Press, Massachusetts, 1977).
42. Oates, J. F. Z. *Tierpsychol.* **45**, 1–160 (1977).
43. Kawabe, M. & Mano, T. *Primates* **13**, 213–227 (1972).
44. Chivers, D. J. in *Primate Conservation* (eds Prince Rainier III & Bourne, G. H.) (Academic, New York, 1977).
45. Kennard, M. A. & Willner, M. D. *Endocrinology* **28**, 967–976 (1941).
46. Wislocki, G. B. *J. Mammal.* **23**, 281–287 (1942).
47. Hall-Craggs, E. C. B. *Proc. zool. Soc. Lond.* **139**, 511–514 (1962).
48. Benoit, J. C. r. *Séanc Soc. Biol.* **87**, 1387–1390 (1922).
49. Daly, M. & Wilson, M. *Sex, Evolution and Behavior* (Duxbury, Massachusetts, 1978).

Relative brain size and basal metabolic rate in terrestrial vertebrates

R. D. Martin

Department of Anthropology, University College, Gower Street, London WC1E 6BT, UK

Studies of the relationship between brain size and body size in terrestrial vertebrates have a long history^{1–4}. Demonstrations of regular relationships between brain and body size across species within selected vertebrate groups serve two purposes: (1) in comparison of species of different body size, empirically recognized 'scaling effects' can be taken into account; (2) empirical relationships may suggest useful working hypotheses regarding functional constraints (although they cannot directly reveal casual connections). It is widely accepted^{5,6} that brain size is scaled to keep pace with changes in body surface area (rather than volume), and this provides the basis for many interpretations of relative brain size. Re-examination of brain-body size relationships for large samples of species from three major vertebrate groups (mammals, birds, reptiles) now shows that there is no empirical foundation for the concept of scaling to body surface area. Instead, it seems that brain size may be linked to maternal metabolic turnover. This has implications not only for assessment of relative brain size in particular species, but also for pursuing links between brain size and 'life strategies'.

A wide variety of studies concerned with scaling of specific parameters to body size have confirmed the general applicability of the allometric formula^{5–8}:

$$Y = kX^a$$

giving the linear logarithmic equation

$$\log Y = a \log X + \log k$$

where Y is the dimension of a parameter such as brain size, X is body size, k is the allometric coefficient and a is the allometric exponent. It is commonly found that different taxonomic groups will exhibit very similar values for a —supporting the interpretation that this reflects fundamental functional relationships—but differ in values for k , because of different 'grades' in development of the parameter (such as brain size) relative to body size⁹.

Probably the best example of graded allometric variation is provided by studies of basal metabolic rate (M) relative to body weight (P). It was originally believed¹⁰ that basal metabolic rate must be related to body surface area because of surface-volume effects in heat exchange

$$M = kP^{2/3} \quad (1)$$

However, rigorous analysis of reliable data¹¹ eventually demonstrated that in mammals the allometric formula actually takes the form

$$M = kP^3 \quad (2)$$

Taking Kleiber's carefully selected data for 26 placental mammal taxa, the major axis¹² gives

$$\log_{10} M = 0.76 \log_{10} P - 0.45 \quad (3)$$

(kcal per day) (g)

($r = 0.998$; 95% confidence limits on $a = 0.74–0.77$). (The major axis is used throughout, in preference to the regression, as it takes into account 'errors' in both X and Y variables. However, use of regressions does not alter any of the basic conclusions. For a discussion of the statistical background, see ref. 12.) The typical exponent value of 0.75 has been abundantly confirmed in other studies, and the following sets of increasing grades in basal metabolic rate relative to body size have been recognized:

- (1) Unicellular organisms: ectothermic vertebrates (fish + amphibians + reptiles): endothermic vertebrates (birds + mammals)^{13,14}.
- (2) Non-passerine birds: passerine birds¹⁵.
- (3) Marsupial mammals: placental mammals¹⁶.

With respect to brain weight (E) and P , the current consensus^{5,6} is that the allometric formula for interspecific relationships within any major vertebrate group takes the form

$$E = kP^{2/3} \quad (4)$$

It has been explicitly suggested that there is a direct link between relative brain size and body surface-volume relationships in vertebrates (for example, because sensors and effectors are often distributed over surfaces)^{4–6}. Indeed, despite variations in shape, for a sample of 37 mammal species^{13,17–22} there is the following relationship between body surface area (S) and P :

$$\log_{10} S = 0.67 \log_{10} P + 1.03 \quad (5)$$

(cm²) (g)

($r = 0.99$; 95% confidence limits for $a = 0.63–0.70$). However, previous analyses of brain size allometry in the major vertebrate groups have been limited by small sample sizes and/or inadequate testing of the value for the allometric exponent. Re-analysis of brain size allometry in placental mammals, taking a representative sample of 309 species from 13 orders^{23–27}, yields the following allometric formula (Fig. 1):

$$\log_{10} E_M = 0.76 \log_{10} P + 1.77 \quad (6)$$

(mg) (g)

($r = 0.96$; 95% confidence limits for $a = 0.73-0.78$). Hence, for placental mammals there is no empirical justification for the widely accepted value of 0.67 for the allometric exponent. The value of 0.76 (equation (6)), confirmed by at least two independent analyses of large samples of mammal species^{6,28}, suggests instead some connection between brain size and basal metabolic rate in mammals. (The reported²⁸ exponent value of 0.73 has been further confirmed by J. F. Eisenberg (personal communication).)

The commonly accepted relationship between brain size and body surface area can be further tested with other vertebrate groups. A sample of 180 bird species^{29,30} yields the following allometric formula (Fig. 2)

$$\log_{10} E_B = 0.58 \log_{10} P + 2.11 \quad (7)$$

(mg) (g)

($r = 0.93$; 95% confidence limits for $a = 0.54-0.61$). A sample of 59 reptile species^{29,31} gives (Fig. 2).

$$\log_{10} E_R = 0.54 \log_{10} W + 1.22 \quad (8)$$

(mg) (g)

($r = 0.96$; 95% confidence limits for $a = 0.50-0.58$). In both cases, the 95% confidence limits of the major axes exclude the hypothesis that brain size is directly related to body surface area (requiring $a = 0.67$), although the exponent values here are significantly below expectation. Thus, it turns out that there is no empirical justification for an exponent value of 0.67 for brain-body size relationships within any of the three major terrestrial vertebrate groups.

Obviously, because both reptiles and birds differ from placental mammals in exhibiting allometric exponents considerably below $a = 0.75$ for brain-body size relations, there is no case for suggesting a direct link between the brain size of the adult and its own basal metabolic rate. However, as mammals are viviparous, whereas birds and most reptiles are oviparous, it may be the development of the brain which relates brain size to the basal metabolic rate of the mother in mammals. (Ideally, one might relate embryonic brain growth to the mother's total metabolic turnover, but insufficient data are available to do so as yet. However, note that embryonic brain growth might be linked to the mother's minimal metabolic

turnover during the day (that is, to her basal metabolic rate). From equation (2), it follows that

$$M_M = k_1 P^{\frac{2}{3}} \quad (9)$$

where M_M is basal metabolic rate of the mother and P is her body weight. If the development of the embryonic brain depends on maternal metabolic turnover, neonatal brain weight (E_N) could show the following relationship.

$$E_N = k_2 M_M = k_2 k_1 P^{\frac{2}{3}} \quad (10)$$

Further, because brain tissue growth is generally restricted to expansion of existing neurones and addition of glial cells following birth, one might expect that adult brain weight (E_A) would be directly proportional to (that is, isometric with) neonatal brain weight

$$E_A = k_3 E_N \quad (11)$$

Equations (10) and (11) combined yield the overall relationship

$$E_A = k_4 P^{\frac{2}{3}} \quad (12)$$

This has already been empirically determined (equation (6) and Fig. 1).

Testing of equations (10) and (11) is complicated by a dichotomy between placental mammals which produce poorly developed, small-brained altricial neonates after a relatively short gestation period, and those which produce well developed, large-brained precocial neonates after a relatively long gestation period³²⁻³⁴. As expected, these two groups of mammals represent two grades of brain evolution with similar exponent values close to 0.75. For precocial mammals ($n = 159$)

$$\log_{10} E = 0.72 \log_{10} P + 2.03 \quad (13)$$

(mg) (g)

($r = 0.93$; 95% confidence limits for $a = 0.67-0.76$) while for altricial mammals ($n = 87$)

$$\log_{10} E = 0.79 \log_{10} P + 1.62 \quad (14)$$

(mg) (g)

($r = 0.96$; 95% confidence limits for $a = 0.74-0.84$) The limited data so far available for mammalian neonatal brain

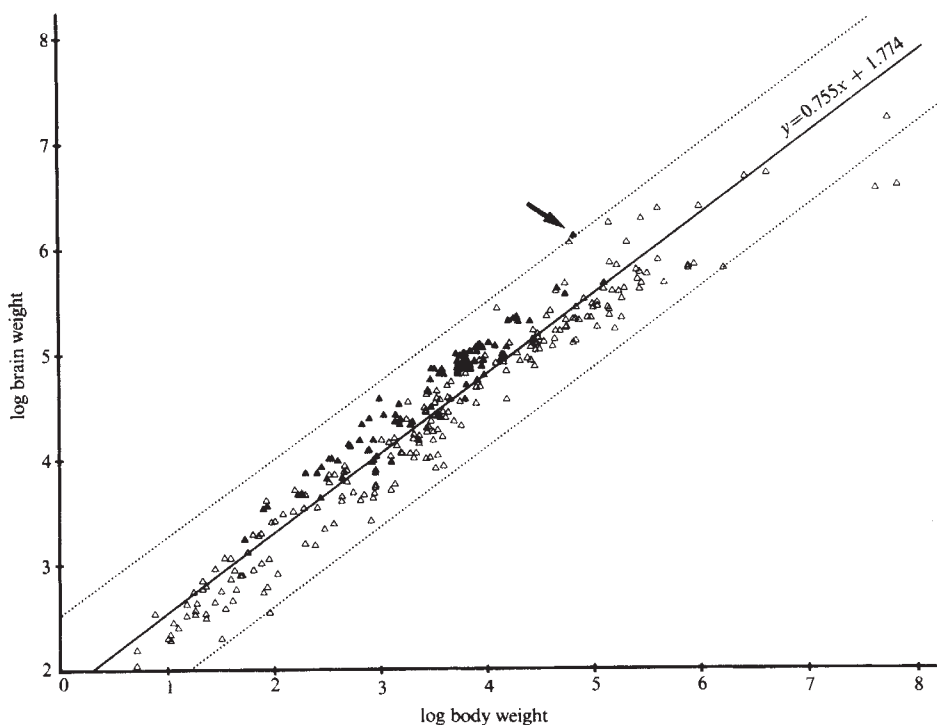
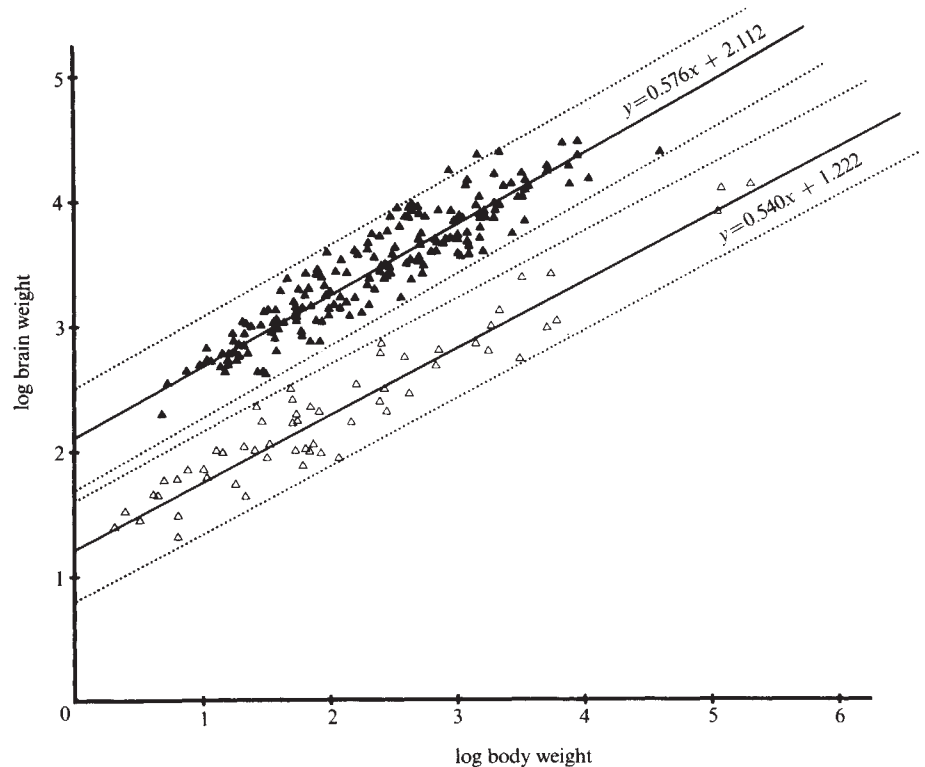


Fig. 1 Allometric relationships between brain and body weights for 309 extant placental mammal species. The solid line is the major axis; dotted lines represent fivefold variation above and below. Primates ($\blacktriangle$) are typical of precocial mammals in exhibiting relatively large brains. The closest species to man (arrowed) is not another primate but a cetacean (dolphin).

Fig. 2 Allometric relationships between brain and body weights for 180 extant bird species ($\blacktriangle$) and 59 extant reptile species ($\triangle$). Solid lines are major axes; dotted lines represent 2.5-fold variation above and below. There is no overlap between the birds and reptiles and the separation between the major axes correspond to a typical 10-fold difference in brain size between birds and reptiles (see ref. 7 for an example of grade differentiation).



weights^{35,36} yield equations which are in good agreement both with equation (10):

Precocial mammals ($n = 71$)

$$\log_{10} E_N = 0.70 \log_{10} P + 1.65 \quad (15)$$

(mg) (g)

($r = 0.94$; 95% confidence limits on $a = 0.64 - 0.76$).

Altricial mammals ($n = 24$)

$$\log_{10} E_N = 0.74 \log_{10} P + 0.88 \quad (16)$$

(mg) (g)

($r = 0.95$; 95% confidence limits on $a = 0.64 - 0.84$) and with equation (11):

Precocial mammals ($n = 71$)

$$\log_{10} E_A = 0.99 \log_{10} E_N + 0.42 \quad (17)$$

(mg) (mg)

($r = 0.99$; 95% confidence limits on $a = 0.96 - 1.02$).

Altricial mammals ($n = 24$)

$$\log_{10} E_A = 1.01 \log_{10} E_N + 0.85 \quad (18)$$

(mg) (mg)

($r = 0.98$; 95% confidence limits on $a = 0.92 - 1.10$).

Whereas in a female placental mammal nutrients are transferred directly to the fetus, thus linking maternal metabolic turnover directly to development of the fetal brain, in birds and most reptiles the mother must first lay an egg and the metabolism of the egg then provides for development of the embryonic brain. There are thus two intermediate equations which may be postulated, expanding on the set listed for mammals

$$M_M = k_1 P^{\frac{2}{3}} \quad (9)$$

$$W = k_5 M_M = k_5 k_1 P^{\frac{2}{3}} \quad (19)$$

$$M_E = k_6 W^{\frac{2}{3}} \quad (20)$$

$$E_H = k_7 M_E \quad (\text{see equation (10)})$$

$$E_A = k_8 E_H \quad (\text{see equation (11)})$$

from which

$$E_A = k_9 (P^{\frac{2}{3}})^{\frac{3}{2}} = k_9 P^{0.56} \quad (21)$$

where W is egg weight, M_E is metabolic rate of egg and E_H is hatchling brain weight, equivalent to E_N for mammals.

The theoretically expected exponent value of 0.56 is compatible with the empirical values determined in equations (7) and (8). Equation (19) has been tested for a sample of egg weights from 127 bird species³⁷, yielding the allometric formula.

$$\log_{10} W = 0.79 \log_{10} P + 2.34 \quad (22)$$

(mg) (g)

($r = 0.95$; 95% confidence limits for $a = 0.75 - 0.84$).

Further, for a considerably smaller sample of 10 bird species, the major axis indicates an allometric exponent of $a = 0.75$ (95% confidence limits = 0.66 - 0.85) for the relationship between metabolic rate and egg weight ($r = 0.98$)³⁸, thus providing provisional confirmation for equation (20). Insufficient data are available to test equations (10) and (11) for birds and reptiles or equations (19) and (20) for reptiles but all the other predictions arising from the hypothesis linking adult brain weight to metabolic turnover in the three major vertebrate groups have been satisfactorily fulfilled.

It is obvious that any association between basal metabolic rate and adult brain size must be seen in terms of general relationship. For placental mammals (Fig. 1), individual values for brain size may vary by a factor of 5 on either side of the value for any given body weight indicated by the major axis, and this variation is not matched by equivalent variation in basal metabolic rate, so different grades of relative brain size among mammals must be influenced by other factors. One obvious factor is the gestation period (relative to body size), as has already been indicated in equations (13) and (14) (see Fig. 1).

Further, it is well established that the value of the allometric exponent for brain-body size relationships in placental mammals decreases with decreasing rank of the taxonomic category studied³⁹. In the absence of any theoretical understanding of this phenomenon, some justification must be given for the groups selected for analysis. Analysis of data for the classes Mammalia, Aves and Reptilia can be justified on several grounds. First, these are taxa of equivalent rank and have long been recognized as distinct grades of vertebrate evolution. Second, taxa of this rank are required to provide adequate sample sizes and a full range of body sizes. Finally, the three taxa are appropriate in

energetic terms, as reptiles have remained ectothermic while birds and mammals have independently developed endothermy.

With these two provisos, it can be concluded from the above empirical relationships that in placental mammals, birds and reptiles relative brain size may be constrained both by the resources channelled to the embryo from the mother and by mode of reproduction (vivipary versus ovipary).

In functional terms, confirmation of this hypothesis would have important implications for research in several areas. There is already strong evidence that the brain is intimately involved in the integration of reproductive parameters^{34-36,40-42} which are themselves intimately interconnected^{36,42-45}. Indeed, it seems likely that the fetal brain may act as a pacemaker in fetal development³⁵, while the size of the adult brain may be associated with ecological aspects such as natural foraging behaviour^{28,42,46}. Hence, further elucidation of possible links between metabolic turnover and brain size could make a major contribution to emerging synthetic models of the organization of the life-history strategies in vertebrate species.

Whether or not the above interpretations survive further testing, the empirical relationships relating brain size to body size in mammals, birds and reptiles have immediate and lasting practical applications for the 'scaling' of brain size in these vertebrate groups. First, Jerison's use⁵ of an allometric exponent value of $a = 0.67$ in his calculation of 'encephalization quotients' for individual placental mammal species needs re-examination. If the correct value of the exponent is close to 0.75, as suggested here for placental mammals (equation (6)), then Jerison's procedure overestimates 'expected' brain sizes for small-bodied mammals and underestimates them for large-bodied mammals. Secondly, if an allometric exponent of 0.56, rather than 0.67, is appropriate for both birds and reptiles (equations (7) and (8)), the converse applies and values for 'expected' brain sizes in large-bodied fossil forms must be revised downwards following the empirical allometric formulae. This is directly relevant to discussion of relative brain size in large-bodied dinosaurs^{5,47-49}, as the new empirical allometric formula established above for modern reptiles indicates that in some species brain size was in fact significantly larger than expected. In modern reptiles, actual values for brain size in individual species may depart from the expected value by a factor of 2.5 on either side of the major axis, so a range of brain indices from 0.4 to 2.5 can be taken as the expected range of natural variation. Appropriately revised values for brain indices in fossil reptiles using available data^{5,47-51}, fall within this range in most cases (sauropods, ornithomorphs, ankylosaurs, stegosaurs, ceratopsians, pterosaurs and cynodont therapsids). However, in two saurischian dinosaurs (the carnosaur *Allosaurus* and the coelurosaur *Stenonychosaurus*) and in the protobird *Archaeopteryx*, brain indices above the expected upper limits are found. Although there is no overlap between modern endothermic birds and ectothermic reptiles in the logarithmic plot of brain and body sizes (Fig. 2), brain indices calculated relative to the modern bird major axis for *Archaeopteryx* (0.43) and for *Stenonychosaurus* (0.60) do lie within the range for modern birds (0.4-2.5). Thus, if there is a connection between relative brain size and metabolic turnover, there are indications that certain dinosaurs, at least, exhibited an advance beyond the modern reptilian condition in metabolic terms.

The ideas developed here owe much to discussions with Dr B. C. C. Rudder and Ms L. Aiello, while considerable assistance was given in data collection, analysis and interpretation by Ms A. MacLarnon. Advice on certain aspects was given by Drs M. Hills, A. Milner and P. H. Harvey.

Received 10 November 1980; accepted 17 July 1981.

1. Snell, O. *Arch. Psychiat. Nervkrankh.* **23**, 436-446 (1891).
2. Dubois, E. *Bull. Soc. Anthropol. Paris* **8**, 337-376 (1897).
3. Lapicque, L. C. *r. Séanc. Soc. Biol.* **5**, 10 (1898).
4. Brandt, A. *Bull. Soc. imp. nat. Moscou* **40**, 525-543 (1867).
5. Jerison, H. J. *Evolution of the Brain and Intelligence* (Academic, New York, 1973).
6. Bauchot, R. *Brain Behav. Evol.* **15**, 1-18 (1978).
7. Huxley, J. S. *Problems of Relative Growth* (Methuen, London, 1932).
8. Gould, S. J. *Biol. Rev.* **41**, 587-640 (1966).
9. Martin, R. D. *Z. Morph. Anthropol.* **71**, 115-124 (1980).
10. Rubner, M. Z. *Biol.* **19**, 535-562 (1883).

11. Kleiber, M. *The Fire of Life: An Introduction to Animal Energetics* (Wiley, New York, 1961).
12. Harvey, P. H. & Mace, G. M. in *Current Problems in Sociobiology* (Cambridge University Press, in the press).
13. Hemmingsen, A. M. *Rep. Steno Meml Hosp.* **4**, 7-58 (1950).
14. Hemmingsen, A. M. *Rep. Steno Meml Hosp.* **9**, 1-110 (1960).
15. Lasiewski, R. C. & Dawson, W. R. *Condor* **69**, 13-23 (1967).
16. Dawson, T. J. & Hulbert, A. J. *Am. J. Physiol.* **218**, 1233-1238 (1970).
17. Liu, C. T. & Higbee, G. A. *J. appl. Physiol.* **40**, 101-104 (1976).
18. Spector, W. S. *Handbook of Biological Data* (Saunders, Philadelphia, 1956).
19. Adams, T. in *Comparative Physiology of Thermoregulation* Vol. 2 (ed. Whitton, G. C.) 151-189 (Academic, New York, 1971).
20. Jarman, P. J. *J. Zool. Lond.* **166**, 349-356 (1972).
21. Ghubrial, L. I. *Physiol. Zool.* **43**, 249-256 (1970).
22. Brody, S. *Bioenergetics and Growth* (Rheinhold, New York, 1945).
23. Bauchot, R. & Stephen, H. *Mammalia* **30**, 160-196 (1966); **33**, 225-275 (1969).
24. Crile, G. W. & Quiring, D. P. *Ohio J. Sci.* **40**, 219-259 (1940).
25. von Bonin, G. *J. gen. Physiol.* **16**, 379-389 (1937).
26. Count, E. W. *Ann. N. Y. Acad. Sci.* **46**, 993-1122 (1947).
27. Stephen, H., Bauchot, R. & Andy, O. J. in *The Primate Brain* (eds Nobak, C. R. & Montagna, W.) 289-297 (Appleton, New York, 1970).
28. Eisenberg, J. F. & Wilson, D. E. *Evolution* **32**, 740-751 (1978).
29. Crile, G. W. & Quiring, D. P. *Ohio J. Sci.* **40**, 219-259 (1940).
30. Portmann, A. *Alauda* **14**, 2-20 (1946); **15**, 1-15 (1947).
31. Platel, R. in *Biology of the Reptilia* Vol. 9 (ed. Gans, C.) 147-171 (Academic, London, 1979).
32. Portmann, A. *Biomorphology* **1**, 109-126 (1939); *Rev. suisse zool.* **72**, 668-666 (1965).
33. Wirz, K. *Acta anat.* **63**, 449-508 (1966).
34. Martin, R. D. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 265-297 (Plenum, New York, 1975).
35. Sacher, G. A. & Staffeldt, E. F. *Am. Nat.* **108**, 593-616 (1974).
36. Rudder, B. C. C. thesis, Univ. London (1979).
37. Heinroth, O. in *Tabulae Biologicae* Vol. 5 (eds Oppenheimer, C. & Pincussen, L.) 716-741 (Junk, Berlin, 1930).
38. Rahn, H., Paganelli, C. V. & Ar, A. *Resp. Physiol.* **22**, 297-309 (1974).
39. Gould, S. J. *Contr. Primatol.* **5**, 244-292 (1975).
40. Sacher, G. A. in *CIBA Fdn Colloq. Ageing* Vol. 5 (eds Wolstenholme, G. E. W. & O'Connor, M.) 115-133 (Churchill, London, 1959).
41. Müller, F. *Verh. naturf. Ges. Basel* **80**, 1-31 (1969).
42. Western, D. *Afr. J. Ecol.* **17**, 185-204 (1979).
43. Payne, P. R. & Wheeler, E. F. *Nature* **215**, 849-850, 1134-1136 (1967); *Proc. Nutr. Soc.* **27**, 129-138 (1968).
44. Leutenegger, W. *Folia primatol.* **20**, 280-293 (1973).
45. Robbins, C. T. & Robbins, B. L. *Am. Nat.* **114**, 101-116 (1979).
46. Clutton-Brock, T. H. & Harvey, P. H. *J. Zool. Lond.* **190**, 309-323 (1980).
47. Bakker, R. T. *Evolution* **27**, 636-658 (1971); *Nature* **238**, 81-85 (1972).
48. Hopson, J. A. A. *Rev. Ecol. Syst.* **8**, 429-448 (1977); in *Biology of the Reptilia* Vol. 9 (ed. Gans, C.) 39-146 (Academic, London, 1979).
49. Bennett, A. F. & Dalzell, B. *Evolution* **27**, 170-174 (1973).
50. Colbert, E. H. *Am. Mus. Novit.* **2076**, 1-16 (1962).
51. Hopson, J. A. in *A Cold Look at the Warm-blooded Dinosaurs* (eds Thomas, R. D. K. & Olson, E. C.) 287-310 (AAAS, Washington, 1980).

Selective loss of binocular depth perception after ablation of cat visual cortex

M. Kaye, D. E. Mitchell & M. Cynader

Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

The massive visual projection to cortical areas 17 and 18 (striate and parastriate cortex) of the cat strongly suggests that these structures have a major role in visual processing. It is therefore paradoxical that ablation of these areas has been reported to result in very trivial deficits in their visual perception and behaviour^{1,2}. In contrast to humans and monkeys, which show profound visual deficits after ablation of the visual cortex³⁻⁶, cats with similar lesions show essentially normal visual behaviour and are able to discriminate complex visual patterns. The only apparent effect of cortical lesions is a mild impairment of the visual acuity for gratings and a possibly greater reduction of vernier acuity⁷. In the past, the functional deficits resulting from lesions or ablations of various cortical structures have provided important clues to the function of many cortical areas, but recently a more detailed picture has emerged from electrophysiological studies of the properties of their constituent neurones. A striking aspect of most visual cortical cells in both cat and monkey is their specificity for horizontal retinal disparity⁸⁻¹⁰, the detection of which is a prerequisite for stereopsis¹¹. This suggests that the visual cortex may play a major part in mediating stereopsis. We report here that ablation of cortical areas 17 and 18 in the cat results in severe selective deficits in binocular depth perception consistent with a loss of stereopsis.

Two adult cats were trained to make simple depth discriminations on a jumping stand illustrated schematically in Fig. 1a. The procedure, described in detail elsewhere¹², consisted of rewarding cats with food for jumping towards the closer of two transparent surfaces covered with randomly placed dots of various sizes; masks immediately below the glass top of the jumping stand hid the edges of the surfaces. The separation of the stimuli, ΔD , was progressively decreased in small steps between blocks of 5 or 10 trials until a threshold separation was reached at which the animal could no longer achieve criterion performance (4 out of 5, or 7 out of 10 correct).

Binocularly, both animals were quickly able to discriminate separations (ΔD) of only 2 cm from a distance of 75 cm, equivalent to a retinal disparity of <4 arc min. Immediately after the animals had attained binocular performance at about this level, they were tested monocularly with one eye covered by an opaque contact lens. The difference in performance in initial trials was striking. Both animals experienced great difficulty, responding at chance on the first session for at least one of the two eyes, even when the two stimuli were separated by 23 cm and viewed from only 35 cm. Even after several training sessions the monocular performance was poorer by a factor of about eight than the best binocular performance.

The obvious difficulty the animals experienced monocularly implies that they had to learn a different discrimination, possibly based on differences in texture size, or even motion parallax, although we observed no obvious increase in head movement when one eye was occluded. The left-hand side of Fig. 2 shows, for one of the two animals (K 215), monocular and binocular thresholds obtained on several consecutive sessions beginning a few sessions before the start of monocular testing. This clear superiority of binocular depth perception is exactly that expected from animals with stereopsis, a conclusion supported by studies of the depth perception of cats subjected to forms of early visual deprivation known to produce a drastic reduction in the proportion of binocular cells in areas 17 and 18 (refs. 13, 14). Earlier studies have shown that such animals lack stereopsis¹⁵ and moreover exhibit no binocular advantage in simple depth

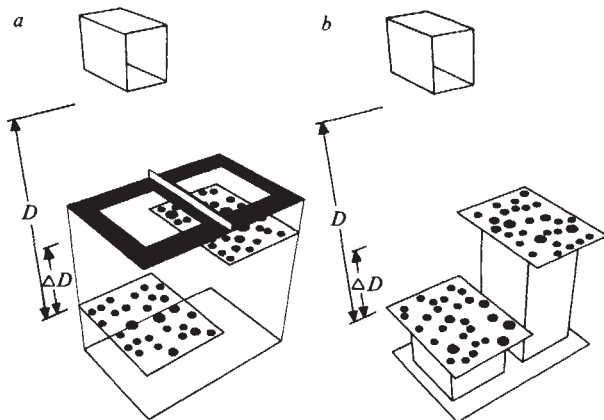


Fig. 1 Two tests of the depth perception of cats. *a*, Animals were trained to leap towards the closer of two transparent surfaces covered with dots of size 0.9–1.9 cm with 18% coverage and illuminated from below. Masks immediately underneath the glass top of the jumping stand on which the cats landed, hid the edges of the stimuli from view. The stimuli had a luminance of ~ 300 cd m⁻². The apparatus eliminated cues provided by the edges of stimuli and by differences in the illumination of the surfaces. The major monocular cues remaining are differences in the relative size and density of the dots. *b*, A simpler test of depth perception. The cats were required to jump from an open-ended box on to the closer of two surfaces covered with dots of various sizes. The separation of the two surfaces, ΔD , was progressively decreased between blocks of trials until the animals were unable to achieve criterion performance (at least 7 correct in a block of 10 trials). This simple apparatus provides a rich variety of monocular depth cues.

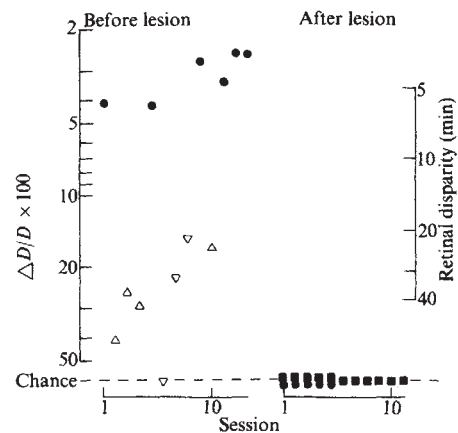


Fig. 2 The performance of K 215 (filled symbols, binocular; open symbols, monocular) before the cortical lesion (left), and both animals afterwards on the apparatus of Fig. 1a (right). The results shown on the left were obtained on several consecutive sessions beginning immediately before the first monocular tests (right eye, upright triangles; left eye, inverted triangles). The left-hand ordinate shows the ratio of the separation of the test surfaces (ΔD) at threshold to the observation distance, D , the distance of the cat from the more distal surface before jumping (corrected for any lean). The right-hand ordinate, which applies only to binocular performance, shows the retinal disparity corresponding to the separation of the two surfaces at threshold. The binocular performance of both K 192 (■) and K 215 (●) following the cortical lesion is shown on the right. Both animals performed at chance.

discriminations¹⁶. The latter finding was confirmed for our testing procedure in an earlier study¹⁷ on a cat reared with a surgically induced divergent strabismus which had almost disappeared at the time of behavioural testing. The binocular and monocular depth thresholds of this animal were virtually identical and comparable with the monocular performance of the two animals of this study. Taken together, these results conclusively show that the lack of binocular superiority in our test indicates a loss of stereopsis.

Cortical areas 17 and 18 and part of area 19 were removed bilaterally from both animals when 15 months old using standard procedures¹⁸. After 14 months, detailed histological reconstruction of the lesions revealed that they were almost complete, except for some sparing on the extreme posterior convexity of the marginal gyrus, on which the extreme periphery (from $\sim 40^\circ$) of the upper visual field is represented.

After several months, when the animals had regained good visuomotor behaviour, their depth perception was again tested on the jumping stand. At about the same time the ability of the animals to resolve gratings was also tested on a jumping stand using procedures described elsewhere¹⁹. Although the latter measurements were not exhaustive, one of the animals (K 215) was able to maintain criterion performance with gratings having spatial frequencies as high as 3.0 cycles deg⁻¹. Less extensive measurements on the other animal (K 192) established that it was able to resolve gratings of at least 2.3 cycles deg⁻¹. Despite the recovery of moderate visual acuity, neither animal was able to solve even the simplest depth discrimination on the jumping stand of Fig. 1a. As indicated in Fig. 2, the animals performed at chance for many sessions even when the two surfaces were separated by 23 cm and viewed from 30 cm. Consequently, they were then tested on the extremely simple apparatus shown in Fig. 1b, which we have previously used¹² for initial training before testing animals on the other apparatus (Fig. 1a). The animals were required to leap directly on to the closer of two adjacent large (35 cm²) surfaces of different heights. Both animals were able to attain criterion performance in this test; Fig. 3 shows the results of all measurements made for both animals. The separation of the two surfaces, ΔD , was kept constant at 8 cm and the observation distance, D , gradually increased

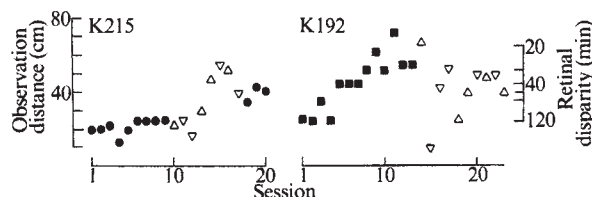


Fig. 3 Binocular (filled symbols) and monocular (open symbols) post-operative performance of K 215 (left) and K 192 (right) on the simple apparatus of Fig. 1b. The separation of the two surfaces, ΔD , was maintained at 8 cm and the observation distance, D , increased until the animals were unable to maintain criterion performance. Monocular symbols are as for Fig. 2.

Conditions required for the inhibitory feedback loop in noradrenergic transmission

D. F. Story, M. W. McCulloch, M. J. Rand & C. A. Stanford-Starr

Department of Pharmacology, University of Melbourne, Victoria 3052, Australia

One of the most striking advances in the study of neurotransmission during the past decade has been the recognition that nerve terminals have receptor sites which, when activated, modulate transmitter release¹⁻⁴. The most extensively studied have been the α -adrenoceptors of noradrenergic sympathetic terminals, activation of which inhibits transmitter release. It is widely believed that the prejunctional α -adrenoceptors are activated by transmitter noradrenaline, which results in inhibition of subsequent transmitter release¹⁻⁴, thereby forming an inhibitory feedback system. However, this hypothesis has recently been challenged by Angus and Korner⁷ on the basis of observations made with short trains of stimulation applied to the intramural sympathetic nerves of isolated guinea pig atria. We have, therefore, re-examined the question using the same tissue and applying not only short, but also longer, trains of stimulation. We report here a time delay before the inhibitory feedback effect through prejunctional α -adrenoceptors is manifested; thus the effect will not be observed unless a sufficiently long train of stimulation is applied. Furthermore, there is no pulse-to-pulse inhibitory feedback regulation of transmitter release for trains of stimulation in which the pulse interval is extended to 8 s.

Guinea pig atria were mounted in a 3-ml organ bath and connected to a strain gauge transducer for measurement of force and rate of beating; basal tension was adjusted to ~1 gramme-force. The solution in the organ bath and in the reservoirs supplying it was kept at 37 °C and continuously gassed with 5% CO₂ in O₂. Two parallel platinum wire electrodes, 1 cm apart, were located at each side of the atria for applying field stimulation. After a 15-min equilibration period, the noradrenergic transmitter in the atria was radiolabelled by incubation in [7-³H] (-)-noradrenaline (1.04 μ M, specific activity 3.2 Ci mmol⁻¹; NEN) for 30 min, followed by a 60-min washout with repeated changes of noradrenaline-free solution. After 30 min of washing, a 30-s train of 1-ms field pulses at 1 Hz was applied to assist in clearing the tissue of non-specifically retained radiolabel.

The atria were then subjected to two periods of field stimulation, 30 min apart, with trains of supramaximal (15 V cm⁻¹), unipolar, square wave pulses. Angus and Korner⁷ delivered field pulses to right atrial preparations, beating spontaneously at ~120 beats per min, as single pulses in the refractory period of four successive beats; however, in our experiments the stimulating pulses were not synchronized with the atrial contractions, so constant frequencies of stimulation could be maintained irrespective of the train duration. Field pulses which occurred outside the refractory period produced additional contractions to those occurring spontaneously; the maximum chronotropic and inotropic responses developed after cessation of stimulation and were largely unaffected by the perturbations in rhythm occurring during stimulation, as shown for chronotropic responses in Fig. 2.

The stimulation-induced release of noradrenaline was deduced from the efflux of radioactivity from the atria⁸. Resting and stimulation-induced effluxes were measured in the solutions taken from the organ bath after 30 s contact with the atria. The resting efflux was expressed as the mean content of radioactivity in two successive 30-s collections taken immediately before

between blocks of trials until the animal could no longer maintain criterion performance. Despite the additional cues to depth provided by the edges of the platform and by small differences in illumination, the binocular thresholds of the two animals on this apparatus were very poor, and no different from their monocular thresholds. Thus there was no evidence that the animals possessed even a rudimentary binocular mechanism for depth perception following ablation of their visual cortex. At the completion of the trials shown in Fig. 3, the animals were again tested on the apparatus of Fig. 1a, without success.

The binocular performance of the lesioned animals was considerably poorer than the monocular performance of normal animals¹⁷, which suggests that visual cortical lesions not only eliminate stereopsis, but also seriously impair the ability to use monocular depth cues. Previous qualitative tests of the depth perception of cats with lesions of areas 17 and 18 have produced conflicting results, suggesting either the loss²⁰, or the retention² of depth perception in situations where both monocular and binocular cues were available.

Our results show that after ablation of areas 17 and 18, cats are unable to solve depth discriminations except in situations (Fig. 1b) that provide a rich variety of monocular cues. However, even in these situations the preoperative binocular superiority of the animals' performance was abolished. These results strongly suggest that ablation of the visual cortex eliminates stereopsis. This implies that the parallel visual pathways to other cortical visual areas^{2,21}, which can mediate moderately high visual acuity and form vision after ablation of areas 17 and 18, nevertheless by themselves cannot process retinal disparity. Moreover, the serious impairment of depth perception in general suggests that the remaining visual pathways are very limited in their capacity to evaluate monocular cues to depth.

This research was supported by grant A7660 from NSERC of Canada to D.E.M. and by USPHS grant EY02248 and grants MT-5201 and A9939 from the MRC and NSERC of Canada to M.C. We thank Mr Michael Greene and Robert Baxter for assistance with the behavioural measurements and Sarah Kirby for assistance with the histology.

Received 30 April; accepted 6 July 1981.

- Doty, R. W. *J. comp. Neurol.* **143**, 341-370 (1971).
- Sprague, J. M., Levy, J., DiBerardino, A. & Berlucchi, G. *J. comp. Neurol.* **172**, 441-488 (1977).
- Brindley, G. S., Gautier-Smith, P. C. & Lewin, W. *J. Neurol. Neurosurg. Psychiatr.* **32**, 259-264 (1969).
- Pasik, T. & Pasik, P. *Vision Res.* **11**, Suppl. 3, 419-435 (1971).
- Weiskrantz, L. *Proc. R. Soc. B182*, 427-455 (1972).
- Miller, M., Pasik, P. & Pasik, T. *J. Neurophysiol.* **43**, 1510-1526 (1980).
- Berkely, M. A. & Sprague, J. M. *J. comp. Neurol.* **187**, 689-702 (1979).
- Ferster, D. *J. Physiol., Lond.* **311**, 623-655 (1981).
- Cynader, M. & Regan, D. *J. Physiol., Lond.* **274**, 549-569 (1978).
- Poggio, G. F. & Fischer, B. *J. Neurophysiol.* **40**, 1392-1405 (1977).
- Wheatstone, C. *Phil. Trans. R. Soc.* **128**, 371-394 (1838).
- Mitchell, D. E., Kaye, M. & Timney, B. *Perception* **8**, 389-396 (1979).
- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 1041-1059 (1965).
- Cynader, M., Mustari, M. J. & Gardner, J. C. *Soc. Neurosci. Symp.* **4**, 99-120 (1979).
- Packwood, J. & Gordon, B. *J. Neurophysiol.* **38**, 1485-1499 (1975).
- Blake, R. & Hirsch, H. V. B. *Science* **190**, 1114-1116 (1975).
- Kaye, M., Mitchell, D. E. & Cynader, M. *Dev. Brain Res.* (in the press).
- Berman, N. & Cynader, M. *J. Physiol., Lond.* **224**, 363-389 (1972).
- Mitchell, D. E., Giffin, F. & Timney, B. *Perception* **6**, 181-193 (1977).
- Cornwell, P., Overman, W., Levitsky, C., Shipley, J. & Lezynski, B. *J. comp. Physiol. Psychol.* **90**, 996-1010 (1976).
- Rosenquist, A. C. & Raczowski, D. *Soc. Neurosci. Abstr.* **6**, 480 (1980).

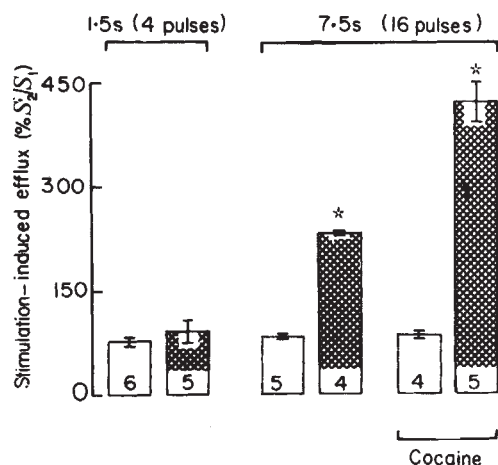


Fig. 1 Effect of phentolamine on the stimulation-induced efflux of radioactivity in guinea pig atria previously incubated with ³H-noradrenaline. The physiological salt solution in the organ bath was a modification of Krebs-Henseleit solution and had the following composition (in mM): NaCl, 118; KCl, 4.7; NaHCO₃, 25; MgSO₄, 0.45; KH₂PO₄, 1.03; CaCl₂, 2.5; D-(+)-glucose, 11.1; disodium EDTA, 0.067; ascorbic acid, 0.15; and atropine sulphate 0.001. Intramural sympathetic nerves were stimulated with field pulses (15 V cm⁻¹, 1 ms, at 2 Hz). Two periods of stimulation of either 1.5 (4 pulses) or 7.5 (16 pulses) s were given with a 30-min interval between them. The stimulation-induced efflux of radioactivity with the second period of stimulation (S₂) was expressed as a percentage of that in the first period (S₁) in control experiments (open columns) and in experiments in which 3 μM phentolamine (hatched columns) was added 20 min before the second period of stimulation and remained present for the second period of stimulation. In some experiments with 7.5 s periods of stimulation, cocaine (30 μM) was added to the solution bathing the atria 15 min before the first period of stimulation and remained present for the second period of stimulation. The columns represent the mean values for the % S₂/S₁; the number of experiments is given at the foot of each column and the s.e.m. is indicated by the vertical line at the top of each column. * Indicates a significant difference between pairs of means ($P < 0.0001$, unpaired 2-tailed, Student's *t*-test). The mean resting effluxes of radioactivity (in 30-s periods) preceding the first period of stimulation in the absence and presence of cocaine were, respectively, 4,147 d.p.m. (s.e.m. = 95, $n = 20$) and 3,951 d.p.m. (s.e.m. = 128, $n = 9$). The absolute values of stimulation-induced efflux with the first period of stimulation were 1,340 d.p.m. (s.e.m. = 139, $n = 11$) with stimulation for 1.5 s and 4,053 (s.e.m. = 250, $n = 9$) and 4,461 (s.e.m. = 475, $n = 9$) d.p.m. with stimulation for 7.5 s in the absence and presence of cocaine, respectively.

stimulation. The stimulation-induced efflux was calculated as the sum of the amounts of radioactivity exceeding the resting level of three successive 30-s collections taken from the beginning of stimulation. In each experiment, the stimulation-induced efflux for the second period of stimulation was expressed as a percentage of that for the first (% S₂/S₁). Contractions of the atria were measured using a semiconductor strain gauge and displayed on a Brush recorder; the signal from the strain gauge was used to trigger a cardiometer coupler, giving a direct recording of rate. The mean resting rate before the first period of stimulation was 194 beats per min (s.e.m. = 4, $n = 59$). Stimulation-induced changes in rate were expressed as for effluxes.

In the first series of experiments, field stimulation was with 1.5-s trains of four pulses at 2 Hz, thus approximating the four-pulse stimulation regime used by Angus and Korner⁷ (see above). Phentolamine (3 μM), added 20 min before the second period of stimulation, had no significant effect on the resting efflux of radioactivity or on the resting force of contractions (paired, 2-tailed, Student's *t*-test; $P > 0.9$); however, in most, but not all, preparations in this and subsequent experiments, the resting rate of beating was depressed by phentolamine; with 1.5-s periods of stimulation the mean decrease was 46 beats per min (s.e.m. = 7, $n = 5$, $P < 0.005$). The stimulation-induced efflux of radioactivity was not significantly affected by phentolamine (% S₂/S₁ approximated control values; Fig. 1); furthermore, the magnitude of the peak increase in rate of beating was not significantly affected (Table 1).

The failure of phentolamine to increase the chronotropic response to stimulation is consistent with the observations of

Angus and Korner⁷, but not with other studies in the same tissue⁹⁻¹¹. The discrepancy may be due to differences in the sympathetic stimulation parameters, in particular the duration of the train of pulses. Stimulation-induced release of ³H-noradrenaline has previously been shown to be enhanced by phentolamine when stimulation was given (either as cardiac accelerans nerve stimulation or as field pulses delivered directly to the atria) with 30- or 60-s trains of pulses at 4 or 5 Hz (refs 9-11) and the peak tachycardia response to sympathetic stimulation was enhanced by phentolamine with accelerans nerve stimulation at 0.5 Hz for 10-s periods¹¹; the increases in each case were attributed to blockade by phentolamine of prejunctional α-adrenoceptors, and hence to disruption of a negative-feedback effect of transmitter noradrenaline, resulting in increased transmitter release. In contrast, our results suggest that there is little or no expression of negative-feedback regulation of transmitter noradrenaline release during neuronal activity evoked by 1.5-s periods of stimulation at 2 Hz.

As Fig. 2 shows, there is a considerable delay before the complete development of the postjunctional responses to sympathetic nerve stimulation. The positive inotropic and chronotropic responses to 1.5-s periods of stimulation developed after stimulation had ended with the peak effect occurring in control preparations 8.7 s (s.e.m. = 0.3, $n = 5$) after the onset of stimulation. The time required for development of the inhibitory feedback action following activation of prejunctional α-adrenoceptors is not known. However, it presumably depends on the times taken for release of the transmitter, the transmitter-receptor interaction and the events leading to the subsequent inhibition of transmitter release.

To investigate whether a delay in the development of negative-feedback action might account for the failure to observe effects of phentolamine when stimulation was for 1.5 s, longer periods of stimulation were used. When the stimulation period was increased to 7.5 s (16 pulses at 2 Hz), phentolamine (3 μM) enhanced the stimulation-induced efflux of radioactivity, the S₂/S₁ ratio increasing ~2.8-fold (Fig. 1). Consistent with the enhanced efflux, there was a significant increase in the peak chronotropic response to stimulation (Table 1, Fig. 2), although the effect was less marked than with efflux, presumably because stimulation with 16 pulses at 2 Hz produced a near-maximal response¹².

These facilitatory effects of phentolamine cannot easily be accounted for in terms of blockade of neuronal re-uptake of transmitter noradrenaline. Thus phentolamine had no such facilitatory effect on efflux or response when stimulation consis-

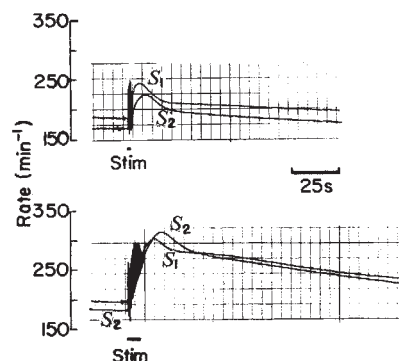


Fig. 2 Effect of phentolamine on the chronotropic responses to field stimulation (Stim) (15 V cm⁻¹, 1 ms, at 2 Hz) of intramural sympathetic nerves in two atrial preparations. In each case the response to the first period of stimulation (S₁) is shown superimposed on the response to the second period (S₂); phentolamine (3 μM) was added 20 min before the second period of stimulation and remained present during that period. In the upper panel the periods of stimulation were 1.5 s (4 pulses) and in the lower panel 7.5 s (16 pulses); bars beneath the records indicate these periods of stimulation. To avoid interrupting the response records, the bathing solution was not collected for radioassay.

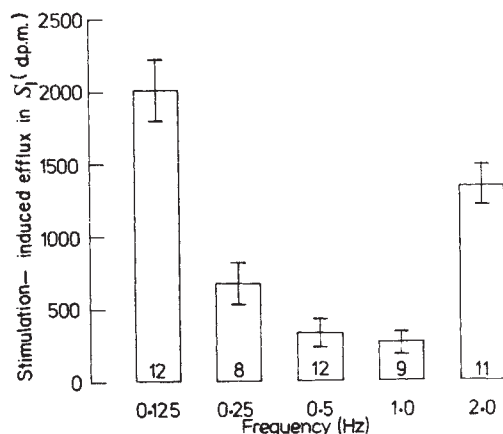
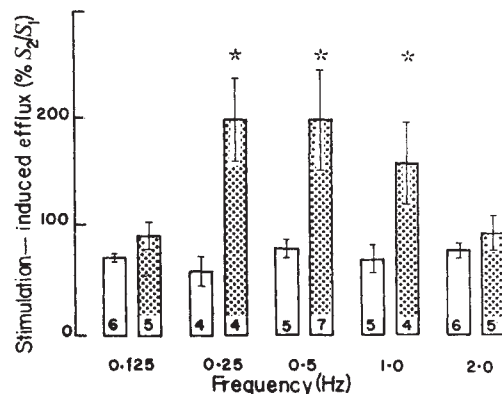
Table 1 Effects of phentolamine on the increases (% S_2/S_1) in rate of beating of guinea pig atria in response to field stimulation of the sympathetic nerves at 2 Hz for periods of 1.5 and 7.5 s

		Stimulation period (s)	
		1.5 (4 pulses)	7.5 (16 pulses)
Control	Mean	108.8	101.4
	s.e.m.	7.0	3.2
	n	6	5
Phentolamine (3 μ M)	Mean	112.0	121.0
	s.e.m.	12.5	6.7
	n	4	5
<i>P</i>		>0.8	<0.03

The peak increase in rate of beating in response to the second (1.5 or 7.5 s) period (S_2) of field stimulation (15 V cm^{-1} , 1 ms at 2 Hz) of intramural sympathetic nerves was expressed as a percentage of the response with the first period (S_1) of stimulation. In some experiments, phentolamine (3 μ M) was added 20 min before the second period of stimulation and remained present throughout that period. Mean values are given together with the standard errors of the means (s.e.m.) and the number of preparations (*n*). *P* values indicate the level of significance for the difference between the means of control and phentolamine groups (unpaired, 2-tailed, Student's *t*-test). The mean increases in rate of beating (Δ beats per min) with the first period of stimulation were 62 (s.e.m. = 11) with stimulation for 1.5 s and 103 (s.e.m. = 9) with stimulation for 7.5 s (*n* = 10 in each case).

ted of four pulses at 2 Hz although the chronotropic responses of guinea pig right atria to field stimulation with four pulses at ~2 Hz are enhanced in the presence of the neuronal uptake-blocking drug desipramine (0.1 μ M)⁷. Furthermore, McCulloch *et al.*¹³ reported that phentolamine in the concentration used in the present study (3 μ M) had no effect on ^3H -noradrenaline uptake by guinea pig atria during incubation in 0.4 μ M ^3H -noradrenaline.

The effect of phentolamine was also determined in the presence of cocaine (30 μ M) using stimulation with 16 pulses at 2 Hz since 30 μ M cocaine has been shown to inhibit the uptake of ^3H -noradrenaline (during incubation in 0.4 μ M ^3H -noradrenaline) by ~70% (ref. 13). Cocaine was added to the solution bathing the atria 15 min before the first period of stimulation and remained present until after the second period of stimulation; in the presence of cocaine, phentolamine (3 μ M) produced even greater enhancement of the stimulation-induced

**Fig. 3** Radioactivity released by first (drug-free) periods of stimulation (S_1) of guinea pig atria with four field pulses (15 V cm^{-1} , 1 ms) delivered at frequencies of 0.125, 0.25, 0.5, 1 and 2 Hz (corresponding to train durations of 24, 12, 6, 3 and 1.5 s). The atria had previously been incubated with ^3H -noradrenaline to radiolabel the noradrenergic transmitter stores. The columns represent the mean values of stimulation-induced efflux, the number of preparations is given at the foot of each column and the s.e.m. is indicated by the vertical line at the top of each column.**Fig. 4** Effect of phentolamine on the release from guinea pig atria of radioactivity evoked by field stimulation with four pulses (15 V cm^{-1} , 1 ms) delivered at frequencies of 0.125, 0.25, 0.5, 1 and 2 Hz (corresponding to train durations of 24, 12, 6, 3 and 1.5 s). The atria had been previously incubated with ^3H -noradrenaline. The stimulation-induced efflux of radioactivity for the second period of stimulation (S_2) was expressed as a percentage of that with the first period (S_1) in control experiments (open columns) and in experiments in which 3 μ M phentolamine was added 20 min before the second period of stimulation (hatched columns). The columns represent the mean values for the % S_2/S_1 ; the number of experiments is given at the foot of each column and the s.e.m. is indicated by the vertical line at the top of each column. * Indicates a significant difference between pairs of means ($P < 0.05$, unpaired, 2-tailed, Student's *t*-test). The mean resting efflux of radioactivity (in 30-s periods) preceding the first period of stimulation was 3,686 d.p.m. (s.e.m. = 107, *n* = 51). The absolute values of stimulation-induced efflux for the first period of stimulation with each of the five regimes are shown in Fig. 3.

efflux of radioactivity than in its absence, the S_2/S_1 ratio increasing ~4.9-fold (Fig. 1). In the presence of cocaine, however, there was no significant enhancement by phentolamine of the chronotropic response to stimulation, presumably because the response was already maximal as a result of uptake blockade. Thus, with the 7.5-s periods of sympathetic stimulation, the findings are entirely consistent with an antagonistic action of phentolamine on prejunctional α -adrenoceptors, thereby disrupting an inhibitory feedback influence on transmitter release which agrees with a previous study⁵ where blockade of prejunctional α -adrenoceptors with phenoxybenzamine failed to enhance transmitter release produced by a single pulse, but did enhance that produced by 16 pulses at 1 Hz. No effects on stimulation with 2, 4 and 8 pulses at 1 Hz were reported⁵, but re-examination of the data shows an excellent correspondence with the present experiments in that transmitter efflux per pulse with two pulses was not significantly enhanced by phenoxybenzamine, but the effluxes with four and eight pulses were.

To investigate further the dependence of α -adrenoceptor-mediated feedback action on the duration of stimulation, the effects of phentolamine were examined on chronotropic responses and stimulation-induced effluxes evoked by four pulses of field stimulation delivered at 1, 0.5, 0.25 and 0.125 Hz (corresponding to train durations of 3, 6, 12 and 24 s). The findings were compared with those for stimulation with four pulses at 2 Hz (1.5-s train duration).

There were marked differences in the amounts of radioactivity in the first (drug-free) periods of stimulation with four pulses at the various frequencies (Fig. 3). The highest release occurred with the lowest frequency of stimulation (0.125 Hz), followed by that with the highest frequency of stimulation (2 Hz); the releases with each of the three intermediate frequencies were similar, but considerably lower than those with either 0.125 or 2 Hz. As shown in Fig. 4, phentolamine (3 μ M) enhanced the efflux of radioactivity evoked by four pulses of

stimulation delivered at frequencies of 1, 0.5 and 0.25 Hz, but had no effect on efflux with stimulation at 2 or 0.125 Hz. Consistent with the effects on transmitter release, the atrial chronotropic responses were unaffected by phentolamine with the highest and lowest frequencies of stimulation, but were enhanced with each of the three intermediate frequencies (Table 2). The failure of phentolamine to enhance efflux or responses, and also the relatively high level of efflux evoked by stimulation with four pulses at both 2 and 0.125 Hz, suggest that in neither conditions of stimulation is there inhibitory feedback regulation of transmitter noradrenaline release. In contrast, the much lower levels of efflux produced by stimulation with four pulses at 1, 0.5 and 0.25 Hz and the enhancements of stimulation-induced effluxes and responses by phentolamine are consistent with inhibitory feedback regulation.

Our findings suggest that inhibitory feedback regulation of transmitter noradrenaline release through prejunctional α -adrenoceptors depends on the frequency and duration of neuronal activity. The failure of phentolamine to enhance transmitter release and the responses evoked by 1.5-s periods of stimulation, in contrast with the enhancements observed with longer periods of stimulation, indicates that for stimulation at a frequency of 2 Hz, manifestation of inhibitory feedback action is delayed at least 1.5 s from the onset of stimulation. Furthermore, the failure of phentolamine to enhance release and responses evoked by stimulation at 0.125 Hz indicates that the duration of activation of the inhibitory feedback effect is less than the pulse interval for this frequency (8 s). It seems likely that the delay in α -adrenoceptor-mediated inhibitory feedback regulation of transmitter release may represent the time required for generation of an intracellular mediator such as a second messenger, and that the persistence of the feedback action may depend on the inactivation of the mediator. Thus the events between prejunctional receptor activation and modulation of transmitter release may closely resemble those linking postjunctional β -adrenoceptor activation with alteration of myocardial activity.

Our results suggest two conditions for the operation of α -adrenoceptor-mediated feedback regulation of sympathetic transmitter release in guinea pig atria, which are consistent with the proposal that transmitter noradrenaline modulates its own release in physiological conditions of cardiac sympathetic nerve activity. The failure of Angus and Korner⁷ to observe potentiation of the tachycardia responses to field stimulation of guinea pig right atria was probably due to their excessively brief periods of stimulation.

This work was supported by grants from the NH and MRC and the National Heart Foundation of Australia.

Table 2 Effects of phentolamine on the increases (% S_2/S_1) in rate of beating of guinea pig atria in response to field stimulation of sympathetic nerves with four pulses at various frequencies

		Four pulses at (Hz)				
		0.125	0.25	0.5	1.0	2.0
Control	Mean	96.8	100.1	99.8	88.8	108.8
	s.e.m.	14.4	10.2	2.8	6.9	7.0
	n	6	5	6	5	6
Phentolamine (3 μ M)	Mean	103.0	174.7	132.7	147.5	112.0
	s.e.m.	9.0	31.0	8.2	9.8	12.5
	n	5	6	6	4	4
P		>0.7	=0.06	<0.005	<0.005	>0.8

The peak increases in rate of beating in response to the second period (S_2) of field stimulation of intramural sympathetic nerves with four pulses (15 V cm^{-1} , 1 ms) at the frequencies indicated are expressed as percentages of the chronotropic response with the first period (S_1) of stimulation. In some experiments, phentolamine (3 μ M) was added 20 min before the second period of stimulation and remained present throughout that period. Mean values are given together with the standard errors of the means (s.e.m.) and the number of preparations (n). P values indicate the level of significance for the difference between the means of the control and phentolamine groups (unpaired, 2-tailed, Student's *t*-test). The mean maximal increases in rate of beating (Δ beats per min) with the first period of stimulation were 28 (s.e.m. = 2, n = 11), 41 (s.e.m. = 8, n = 11), 57 (s.e.m. = 6, n = 12), 54 (s.e.m. = 3, n = 9) and 62 (s.e.m. = 11, n = 10) for stimulation with 0.125, 0.25, 0.5, 1.0 and 2.0 Hz, respectively.

Received 30 September 1980; accepted 16 July 1981.

1. Langer, S. Z. *Br. J. Pharmac.* **60**, 481-497 (1977).
2. Starke, K. *Rev. physiol. biochem. Pharmac.* **77**, 1-124 (1977).
3. Westfall, T. C. *Physiol. Rev.* **57**, 659-728 (1977).
4. Rand, M. J., McCulloch, M. W. & Story, D. F. in *Adrenergic Activators and Inhibitors: Handbook of Experimental Pharmacology* 54/1 (ed. Szekeres, L.) 233-266 (Springer, Berlin, 1980).
5. Rand, M. J., Story, D. F., Allen, G. S., Glover, A. B. & McCulloch, M. W. in *Frontiers in Catecholamine Research* (eds Usdin, E. & Snyder, S. H.) 579-581 (Pergamon, New York, 1973).
6. Rand, M. J., McCulloch, M. W. & Story, D. F. in *Central Actions of Drugs in Blood Pressure Regulation* (eds Davies, D. S. & Reid, J. L.) 94-132 (Pitman, Tunbridge Wells, 1975).
7. Angus, J. A. & Korner, P. I. *Nature* **286**, 288-291 (1980).
8. McCulloch, M. W., Rand, M. J. & Story, D. F. *Clin. exp. Pharmac. Physiol.* **1**, 275-289 (1974).
9. McCulloch, M. W., Rand, M. J. & Story, D. F. *Br. J. Pharmac.* **46**, 523-534P (1972).
10. Adler-Graschinsky, E. & Langer, S. Z. *Br. J. Pharmac.* **53**, 43-50 (1975).
11. Langer, S. Z., Adler-Graschinsky, E. & Giorgi, O. *Nature* **265**, 648-650 (1977).
12. Blinks, J. R. *J. Pharmac. exp. Ther.* **151**, 221-235 (1966).
13. McCulloch, M. W., Medgett, I. C. & Rand, M. J. *Br. J. Pharmac.* **67**, 535-543 (1979).

Selective transfer of Lucifer yellow CH from axoplasm to adaxonal glia

Terry A. Viancour*, George D. Bittner† & Martis L. Ballinger†

*University of Texas Health Science Center, Sensory Sciences Center, Room 7.238, Houston, Texas 77025, USA

†Department of Zoology, University of Texas, Austin, Texas 78712, USA

The role of axon-sheath glia in supplying proteins and other substances to axons is a topic of interest and controversy¹⁻⁹. We have now found that when crayfish giant axons are injected intracellularly with the tracer dye Lucifer yellow CH, within minutes the dye can enter all axon-sheath glia of the adaxonal layer (the glial layer immediately adjacent to the axon). This demonstrates molecular movement from axoplasm to gliaplasm, and satisfies a prerequisite for the communication to glia of specific axonal needs.

Paired medial (MGAs) and lateral (LGAs) giant axons in the crayfish *Procambarus clarkii* are identifiable in each preparation. These axons are surrounded by individual glial sheaths 2-6 μ m thick consisting of alternating layers of glial cytoplasm and extracellular space containing collagen or other material¹⁰⁻¹². The sheath around an MGA is 50-200% thicker than that around an adjacent LGA¹³. The layer of adaxonal glia consists solely of slightly convoluted glial cytoplasm. The adaxonal glia are separated from the axolemma by an extracellular space of 20-80 nm (refs 10, 11). Adaxonal glia cytoplasm contains mitochondria, microtubules and rough endoplasmic

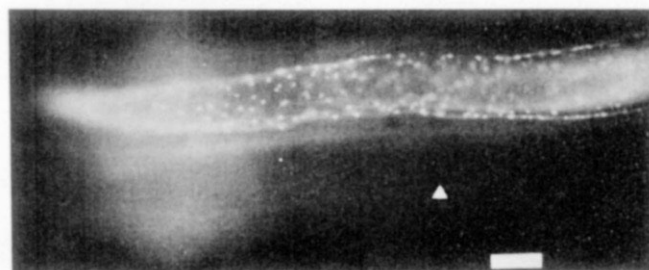


Fig. 1 Fluorescence micrograph of an MGA and the adaxonal glia nuclei of its sheath, near the base of the cerebral ganglion. Rostrally (left), the nuclei on the dorsal surface of the axon are in focus; nuclei on the lateral surfaces are in focus caudally. From such data, we estimate that a medial giant axon in a crayfish 7-10 cm long has about 15,000 adaxonal glia. Note the non-fluorescing image of the adjacent lateral giant axon (arrow). Fluorescent dyes such as Procion or Lucifer often label axolemma and nuclei more intensely than cytoplasm, possibly because of different concentrations of cellular constituents to which the dyes bind. Scale bar, 100 μ m.

reticulum; axoplasm contains neurotubules, mitochondria and smooth endoplasmic reticulum^{10,11}.

The Lucifer yellow CH anion is a lipid-insoluble, sulphonated naphthalimide fluorescent dye with a molecular weight of 443 (ref. 14). Dye was iontophoresed into MGAs or LGAs for 5–60 min, using 20–100-nA pulses of 0.5 s duration at 1 Hz. Immediately after injection, the tissues were fixed overnight in 4% formaldehyde in 0.1 M phosphate buffer (pH 7.4) or 10% formalin and dehydrated in ethanol. Tissues viewed in whole-mount with a transmitted-light fluorescence microscope¹⁴ were cleared with methyl salicylate. To obtain correlative light and electron microscopic data, tissues were embedded in Spurr's plastic and sectioned serially in a sequence of thick and thin sections: 3–5 thick sections (0.5–1 μ m), 6–10 thin sections (60–90 nm), and one final thick section (1–2 μ m). The thin electron microscope sections were collected on single-hole grids, stained with warm 5% uranyl acetate in 95% ethanol and post-stained with lead citrate. Fluorescent structures in the last thick section could be identified unambiguously in electron micrographs of the last thin section.

Figure 1 shows a dye-filled MGA viewed in whole-mount. The axon boundary is defined by a region of slightly more fluorescence and by brightly fluorescing nuclei of adaxonal glia. These nuclei appear as 7- μ m diameter disks lying tangential to the axon surface. They contain one or two brighter 1- μ m spots (visible at higher magnifications) which were identified subsequently as nucleoli. In each of the 30 areas analysed from five MGAs, cells with fluorescing nuclei were always those of adaxonal glia identified by electron microscopy, and nuclei from all such identified adaxonal glia fluoresced (Fig. 2). Nuclei of glia in other sheath layers did not fluoresce above the level observed in uninjected control axons.

Twenty MGAs injected in the circumoesophageal connective and four injected in abdominal connectives all gave comparable results. Injected LGAs showed fluorescent glia patterns similar to the MGAs, except that the number of stained glia per unit length was less than for adjacent MGAs. Stained glia of dye-injected non-giant axons were also observed. Fluorescing glia were not seen where the MGAs are divested of their axon sheath in the supraoesophageal ganglion¹⁵. Mitochondria in dye-filled axons also fluoresced in plastic-embedded sections.

We have considered several unphysiological or nonspecific mechanisms by which the adaxonal glia might have been stained. Passive diffusion into glia is unlikely: Lucifer is very lipid insoluble and membrane impermeable at physiological pH¹⁴; extracellularly iontophoresed dye did not stain glia or axons. If dye entered damaged glia at the site of axon penetration it could have entered other glia through inter-glial¹² gap junctions. We have, however, observed that dye injected near the supraoesophageal ganglion into one MGA enters, presumably via an electrical synapse, the homologous, contralateral MGA and its adaxonal glia. This happens even though the axon sheath is absent and glial nuclei are not stained in the ganglion where the MGAs make their only direct synaptic contact¹⁶. Increased permeability resulting from cell membranes damaged before fixation is also unlikely: Lucifer dye is a relatively nontoxic¹⁴; membranes appeared normal in electron micrographs; injection currents were not unusually high^{17,18}, especially for these large, 100–200- μ m diameter axons; resting membrane potentials remained constant during dye injections; axons were filled in dim light to avoid cytoplasmic damage resulting from intense light exposure of dye-filled cells before fixation¹⁹. Finally, dyes can relocate during histological fixation²⁰, but when an injected, physiologically intact MGA is transected to allow for diffusion and dilution of intra-axonal dye, and viewed immediately, the number and pattern of fluorescent glia are comparable with those observed in fixed tissue.

We suggest that the rapid dye movement into adaxonal glia reflects a special relationship for intercellular molecular exchange between giant axons and glia. Dye diffusion through gap junctions between glia and axons could account for our

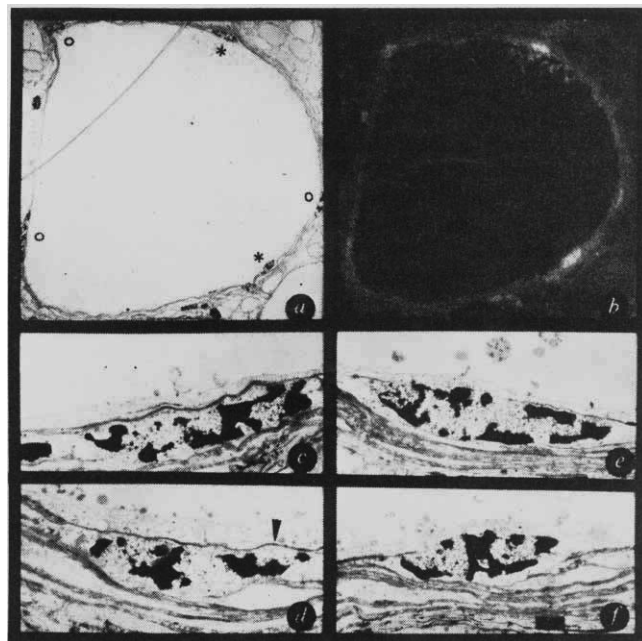


Fig. 2 Micrographs of sequential thick and thin cross-sections of an injected MGA. *a*, Electron micrograph showing glial cell nuclei in the adaxonal (*) and other (O) sheath layers. *b*, Light micrograph showing fluorescing glial cell nuclei (*) in the adaxonal layer. *c*, *d*, Electron micrograph of two of the three glial nuclei in other sheath layers of *a*. Arrow indicates a collagen layer distal to the adaxonal layer (the collagen is not visible in these micrographs). *e*, *f*, Electron micrograph of glial nuclei in adaxonal layer of *a*. Scale bar: *a*, *b*, 10 μ m; *c*–*f*, 1 μ m.

results, although gap-like junctions have not been observed between crustacean glia and axons^{10–12}. Exocytotic–pinocytotic processes could transfer dye^{10,11} and, in Crustacea, regions of specialized axon–glia membrane¹² and vesicle accumulation at sites of glial–neuronal apposition²¹ have been reported. Recently, Peracchia has reported finding areas of apparently unrestricted cytoplasmic continuity between adaxonal glia and axons in crayfish abdominal connectives²².

Whatever the cellular mechanism, it seems that polar substances of at least 443 molecular weight can rapidly and specifically move from a crayfish giant axon into its adaxonal glia. This may be a specific manifestation of a normal process of exchange of many different substances between adaxonal glia and intact or severed axons of crayfish and other organisms^{5–11,13,23}. Indeed, the distal stumps of severed MGAs often survive morphologically intact and functionally competent for 90–300 days. During this time individual layers of the glial sheath, particularly the adaxonal layer, typically increase in thickness^{10,11,24,25}. Glial cells around severed MGAs usually synthesize and transfer more labelled proteins to severed MGAs than to intact MGAs⁹, and adaxonal glial cells synthesize greater amounts of radioactive proteins than do other, non-adaxonal glia in crayfish⁹ and squid^{5–8}.

This work was supported by a PHS traineeship grant to T.A.V., a NIH research grant (NS-14412) and Career Development Award (NS00070) to G.D.B., and a NSF grant (BNS-7910-335) to Raymon Glantz. We thank Dr Walter Stewart of NIH for providing us with Lucifer yellow CH and R. Glantz, M. Kirk and H. Sperling for their contributions.

Received 12 February; accepted 14 July 1981.

- Varon, S. S. & Somjen, G. G. *Neurosci. Res. Prog. Bull.* **17**, 1–239 (1979).
- Kuffler, S. W. & Potter, D. D. *J. Neurophysiol.* **27**, 290–320 (1964).
- Kuffler, S. W., Nicholls, J. G. & Orkand, R. K. *J. Neurophysiol.* **29**, 768–787 (1966).
- Singer, M. & Salpeter, M. M. *J. Morph.* **120**, 281–316 (1966).
- Lasek, R. J., Dabrowski, C. & Nordlander, R. *Nature* **244**, 162–165 (1973).
- Lasek, R. J., Gainer, H. & Przybylski, R. J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1188–1192 (1974).
- Lasek, R. J., Gainer, H. & Barker, J. L. *J. Cell Biol.* **74**, 501–523 (1977).
- Gainer, H., Tasaki, I. & Lasek, R. J. *J. Cell Biol.* **74**, 524–530 (1977).

9. Meyer, M. R. & Bittner, G. D. *Brain Res.* **143**, 212-232 (1978).
10. Ballinger, M. L. & Bittner, G. D. *Cell Tissue Res.* **208**, 123-133 (1980).
11. Bittner, G. D. & Ballinger, M. L. *Cell Tissue Res.* **207**, 143-153 (1980).
12. Peracchia, C. J. *Cell Biol.* **61**, 107-122 (1974).
13. Meyer, M. R. & Bittner, G. D. *Brain Res.* **143**, 195-211 (1978).
14. Stewart, W. W. *Cell* **14**, 741-759 (1978).
15. Bullock, T. H. & Horridge, G. A. *Structure and Function in the Nervous Systems of Invertebrates*, 902 (Freeman, San Francisco, 1965).
16. Viancour, T. A. *Soc. Neurosci. Abstr.* **6**, 575 (1980).
17. Birse, S. C. & Bittner, G. D. *J. Neurophysiol.* **45**, 724-742 (1981).
18. Murphy, A. D. & Kater, S. B. *Brain Res.* **186**, 251-272 (1980).
19. Miller, J. P. & Selverston, A. I. *Science* **206**, 702-704 (1979).
20. Bennett, M. V. L. in *Intracellular Staining in Neurobiology* (eds Kater, S. B. & Nicholson, C.) 115-134 (Springer, New York, 1973).
21. Nordlander, R. H. J. *comp. Neurol.* **161**, 499-514 (1975).
22. Peracchia, C. *Nature* **290**, 597-598 (1981).
23. Bittner, G. D. in *Neurons and Behavior* (ed. Hoyle, G.) 507-532 (Freeman, San Francisco, 1977); *Comp. Biochem. Physiol. A* **68**, 299-306 (1981).
24. Wine, J. J. *Expl Neurol.* **38**, 157-169 (1973).
25. Bittner, G. D., Ballinger, M. & Larimer, J. L. *J. exp. Zool.* **189**, 13-36 (1974).

Antigenic variants of measles virus

Michael J. Birrer, Stephen Udem, Stanley Nathenson & Barry R. Bloom*

Department of Microbiology and Immunology,
Albert Einstein College of Medicine,
1300 Morris Park Avenue, Bronx, New York 10461, USA

Since its first isolation in 1954, measles has been considered to be an antigenically stable virus^{1,2}, a view supported by numerous epidemiological studies demonstrating that a single exposure to measles confers life-long immunity on the infected host³. The basis for protection against recurrent infection is thought to be the production of polyclonal neutralizing antibodies that recognize a virus existing in only one, invariant antigenic form. Indeed, comparison of several independent isolates of measles virus for biological parameters such as cytopathic effect (CPE), host range and haemagglutination has supported the view that variation in measles virus is uncommon^{4,5}. However, there has been no evaluation of variation at the molecular level by rigorous biochemical analysis of measles virus strains and of their structural proteins. We report here the isolation and characterization of spontaneously arising antigenic variants of measles by immunoselection, using neutralizing monoclonal antibodies directed against antigenic determinants of the Edmonston measles haemagglutinin protein. One of these variants has been shown to be antigenically similar to a low-passage 'street' virus isolate of measles, which suggests possible antigenic variation outside the laboratory.

Measles virus belongs to the paramyxovirus family and contains six major structural polypeptides⁶. One of these, the haemagglutinin, of molecular weight (MW) 80,000, is a surface protein which plays an important part in adsorption of the virus to cell surfaces. Presumably, anti-haemagglutinin antibodies neutralize the virus by blocking this function. Monoclonal antibodies to the viral haemagglutinin used here have been described elsewhere⁷. Two monoclonal antibodies (H-1 and F/20) were obtained from high-titre ascites by implantation of hybridoma cells into the peritoneal cavity of pristane-treated mice. Each was capable of neutralizing >200,000 plaque-forming units (PFU) per ml ascites. The specificity for the haemagglutinin polypeptide of measles virus was established by immunoprecipitation as shown in Fig. 1.

Selection for spontaneously occurring antigenic variants of measles virus was performed as follows: 10⁴ PFU of a fifth passage stock of attenuated Edmonston strain of measles virus were mixed with a neutralizing excess of monoclonal antibodies in a total volume of 200 µl and allowed to react on ice for 45 min. This mixture was then plated on to a confluent monolayer of

Vero cells and allowed to adsorb at 31 °C for 2 h. After the adsorption period, the monolayer was incubated in 0.75% CMC (carboxymethyl cellulose) in 10% FCS (fetal calf serum)-DME (Dulbecco's modified Eagle's) medium for 5 days at 31 °C. The presence of viral plaques was assessed after staining with neutral red.

Selection in the presence of monoclonal antibody F/20 gave several plaques, but there was none in the presence of H-1. Fourteen F/20-resistant plaques were transferred to Vero monolayers and passaged to high titre. Neutralization assays were performed to establish that these were antigenic variants of the original stock virus (Table 1). Of the three clones analysed, none was neutralized by the F/20 monoclonal antibody (index of 1), whereas the wild-type stock had a neutralization index of 10⁻⁴. Interestingly, these F/20 variants were neutralized by monoclonal antibody H-1 to the same degree as the wild-type stock. They thus seem to differ from wild type in at least one antigenic determinant that is recognized by the F/20 monoclonal antibody, but share the determinant that is recognized on wild-type virus by the H-1 antibody. To establish that these were stable genetic mutants, the variants were passaged three times more and assayed by neutralization after each passage. In all cases, the variants showed the same pattern of neutralization, that is, non-reactivity with F/20 antibodies (antibodies used for selection) and complete neutralization by H-1 antibodies (data not shown). It proved difficult to estimate the frequency at which these variants arose (due to the titre of our antiserum) but it was

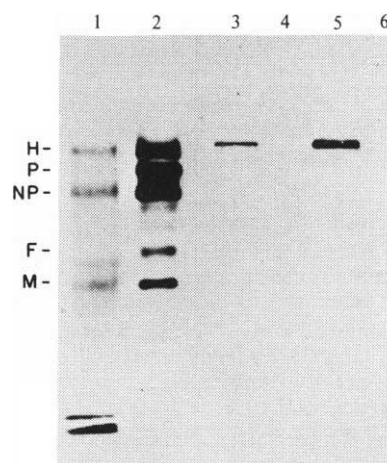


Fig. 1 SDS-PAGE analysis of immunoprecipitated polypeptides from infected cellular extracts. Lane 1, purified measles virus; lane 2, precipitation of infected cellular extracts by hyperimmune rabbit anti-measles antiserum; lanes 3, 5, precipitation of infected cellular extracts by monoclonal antibodies F/20 and H-1, respectively; lanes 4, 6, precipitation of cellular extracts by monoclonal antibodies F/20 and H-1, respectively. H designates the measles haemagglutinin; P, the phosphoprotein; NP, the nucleocapsid protein; F, the fusion protein; and M, the matrix protein. The precipitates were prepared according to ref. 7. Briefly, infected (multiplicity of infection ~0.1) and mock-infected cultures of CV-1 cells were pulse-labelled with either ³⁵S-methionine, ¹⁴C-arginine or ³H-arginine; this was done until 100% CPE at which time the monolayer was washed once with phosphate-buffered saline (PBS) and lysed in a detergent buffer (0.5% NP-40, 0.3% SDS, 0.5% deoxycholate Tris buffer pH 7.4). A 10-µl sample of the extract was precipitated in cold trichloroacetic acid (TCA), washed and counted with Aquasol in a Beckman scintillation counter. 5 × 10⁶ c.p.m. of the extract was added to 20 µl of hyperimmune rabbit anti-measles antiserum or 1 µl of hybridoma ascites fluid. The mixture was allowed to react overnight at 0 °C; then the immune complexes were removed by the addition of protein-A-Sepharose (Pharmacia) and subsequent centrifugation in an Eppendorf microfuge. After several washings, the Sepharose was resuspended in SDS lysis buffer and boiled for 3 min. Aliquots were removed, loaded on to 7.5% SDS-Tris-glycine gels and run for ~2.5 h. The gels were stained, destained and processed for fluorography (Enhance; NEN).

* To whom reprint requests should be addressed.

Table 1 Neutralization analysis of wild-type virus and antigenic variants using monoclonal antibodies

Virus	Virus titre (PFU ml ⁻¹)			No antibody	Neutralization index		
	RαMs	H-1	F/20		RαMs	H-1	F/20
Edmonston	7.5×10^2	7×10^2	1×10^3	7.5×10^6	1×10^4	1×10^4	7.5×10^3
1677	<50	<50	3×10^4	1.5×10^4	6×10^2	6×10^2	0.5
Clone 4	ND	<50	1.5×10^5	1.5×10^5	ND	3×10^3	1.00
Clone 1	ND	<50	1.5×10^5	1.75×10^5	ND	4×10^4	1.09
Clone 2	ND	ND	1.6×10^5	1.8×10^6	ND	ND	11.6

Neutralization assays were performed according to ref. 7. Briefly, 10-fold dilutions of a virus sample were made in the presence and absence of a constant amount of antibody. After 45 min on ice, these dilutions were plated on to a monolayer of Vero cells and allowed to adsorb for 2 h. The cells were then overlaid with 0.75% CMC and incubated at 37 °C for 5 days. Neutralization index = titre of virus sample/titre of virus sample and antibody. ND, not determined. Rα Ms refers to rabbit hyperimmune anti-measles serum; H-1 and F/20 are two monoclonal anti-measles haemagglutinin ascites.

clearly $<10^{-4}$. Biochemical analysis of one variant (clone 1) was performed to establish whether the variant was a true mutant expressing an alteration in the haemagglutinin protein encoded by the virus. Immunoprecipitation with monoclonal antibody H-1 (monoclonal antibody F/20 did not precipitate the clone 1 haemagglutinin) of ³⁵S-methionine-labelled infected extracts

followed by analytical SDS-gel electrophoresis revealed no differences between clone 1 and wild-type haemagglutinins. Isoelectric focusing showed no charge differences between the proteins, each having an isoelectric point of ~pH 6.0, and having the same minor heterogeneous pattern, probably due either to glycosylation or other post-translational modification (data not shown). Immunoprecipitated, preparative gel-purified haemagglutinin, labelled with radioactive arginine or lysine, was then subjected to peptide mapping after tryptic digestion and ion-exchange chromatography (Fig. 2). Co-chromatographic resolution of the ³H-arginine tryptic peptides of the haemagglutinin of clone 1 with those of ¹⁴C-arginine peptides of the Edmonston measles haemagglutinin revealed no differences (Fig. 2a). In contrast, comparison of lysine-labelled peptides showed two significant differences between the two haemagglutinins (Fig. 2b). Two peptides were found in the clone 1 pattern for which there were no corresponding peaks in the wild type. This suggests several possibilities, for example, a new lysine cleavage site or amino acid substitution in two peptides. Note that 33 of the 35 peaks shown in both the arginine and lysine maps are shared between both proteins, which suggests that the two haemagglutinin proteins differ by one or only a very limited number of amino acids. As a control, arginine- and lysine-labelled tryptic peptides of the nucleocapsid proteins were compared for clone 1 and the wild-type virus (Fig. 2c); no differences were detected. Thus we infer that the variants in the haemagglutinin arose spontaneously in tissue culture and were isolated due to the selection of antibody.

Do such antigenic variants exist outside the laboratory? We determined the capacity of the two monoclonal antibodies to neutralize an early-passage measles isolate, the Enders 1677 strain. As shown in Table 1, this 'street virus' showed the same neutralization profile as our variant virus, that is, efficient neutralization by H-1 but not by F/20. As the 1677 strain grows

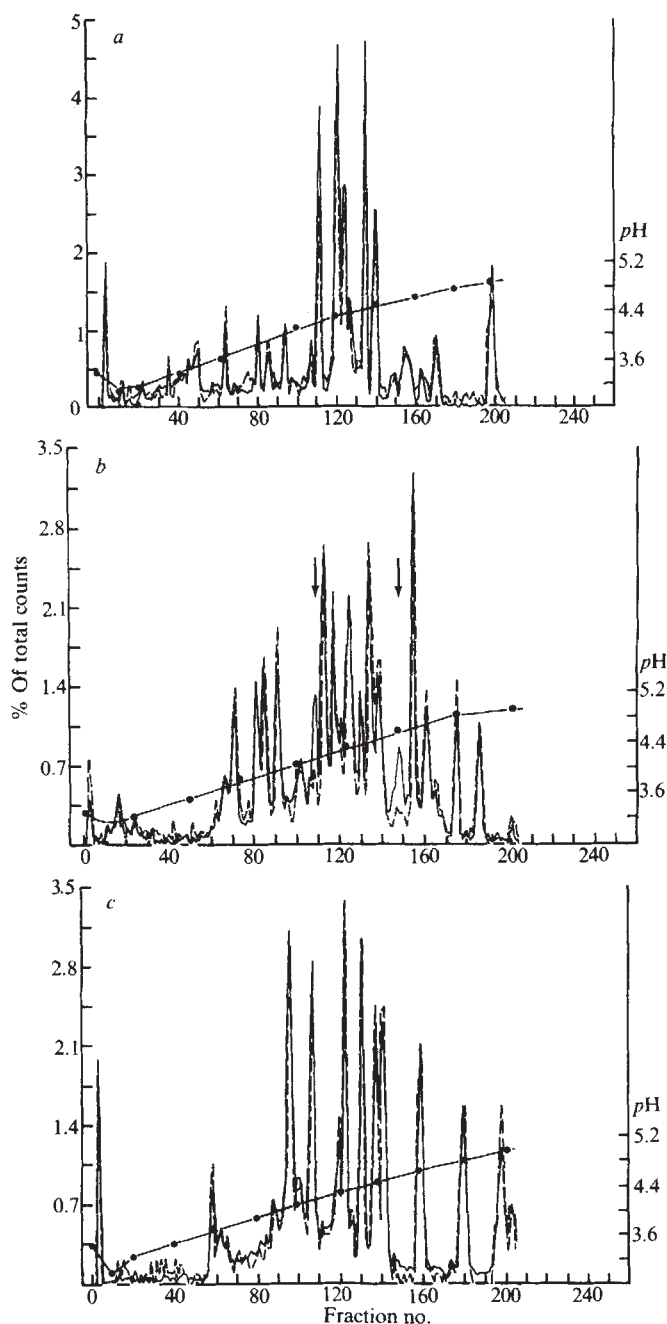


Fig. 2 *a*, Comparison of the tryptic peptides of clone 1 haemagglutinin (³H-arginine) with Edmonston measles haemagglutinin (¹⁴C). *b*, Comparison of the tryptic peptides of clone 1 haemagglutinin (³H-lysine) with Edmonston measles haemagglutinin (¹⁴C). *c*, Comparison of the tryptic peptides of clone 1 nucleocapsid (³H-lysine) with Edmonston nucleocapsid (¹⁴C-lysine). — — —, ¹⁴C data; —, ³H data. The ion-exchange chromatography was performed as described elsewhere¹⁵. Briefly, ³H-lysine-labelled clone 2 haemagglutinin and ¹⁴C-lysine-labelled Edmonston haemagglutinin, immunoprecipitated by monoclonal antibodies and purified from 10% SDS preparative gels, were mixed to give a total of 20,000 and 10,000 c.p.m. of ³H and ¹⁴C, respectively. These proteins were precipitated with 20% TCA, washed, and digested with TPCK trypsin. The supernatant, containing the soluble peptides, was separated on a column of Spherix resin (type XX8-60-0; Phoenix Co.) using a gradient of pyridine-acetic acid buffers. Elution was carried out at 51 °C and 200 fractions of 60 drops per tube were collected in minivials. The pH of the fractions were determined, and after evaporation, they were dissolved in 0.1 ml H₂O and counted with Biophore (NEN) in a LS 7500 scintillation counter (Beckman). Correction was made for the overlap of the ¹⁴C emission spectrum into the ³H channel.

only poorly in tissue culture, it was not possible to perform peptide analysis, and it is likely that the street virus differs from Edmonston virus at more determinants than our *in vitro* selected mutants. Furthermore, note that this study is limited in that we examined only one street-type virus (due to the paucity of early-passage viral isolates). Consequently, it is possible that all or most street viruses are antigenically identical to 1677 and that the Edmonston laboratory strain is the variant. If so, we would then have to reinterpret our *in vitro* selected variants as being revertants back to a viral antigenic type more closely resembling that found in nature. Further examination of other isolates should resolve this.

This is the first clear demonstration of spontaneously occurring antigenic variation of measles virus. However, the following should be considered. The alteration reported here on the haemagglutinin polypeptide occurs at one antigenic determinant selected against by our monoclonal antibody. Variation could be occurring elsewhere in this polypeptide, and even in other proteins of measles virus, but as we have screened using only two monoclonal antibodies, we can only discriminate between variation in two antigenic determinants. Thus, although there is only a low frequency of haemagglutinin variants to F/20 antibody ($< 10^{-4}$), the total frequency of variants in a population could be much greater. Clearly one would expect a number of amino acid or antigenic determinants to be conserved, as has been found for other viral systems⁸. Note that we have been unable to select from mutants using our other monoclonal antibody (H-1), which could be an indication that the antigenic site recognized by this serum is crucial to the integrity of the infectious particle, or that variation is limited to certain 'hot spots' in the gene coding for this molecule.

Our results raise additional questions about the significance of variation in measles virus, which has always been regarded as having a single serotype. The data reported here reveal that variants are produced at about the same rate as those reported for influenza⁹ and other viruses¹⁰. Clearly, measles virus should be re-examined for differences between vaccine strains and various wild-type isolates. More importantly, the present findings indicate that the measles genome is subject to mutation, and the antigens to variation, necessary events to support the view that measles virus mutants are involved in persistent infections^{11,12}. Measles virus mutants have been isolated from subacute sclerosing panencephalitis patients. Thus they can arise in nature^{13,14}, and it should now be possible to determine whether antigenic variants have any role in the persistence of measles-like viruses *in vivo*.

We thank Dr William Hall for the early isolate of 1677 and Dr M. Scharff for helpful suggestions. This work was supported by grants RG 1006 and RG 1241 from the National Multiple Sclerosis Society; AI 09807, AI 10702 and 07289 (to S.N.) from the USPHS, and MV-26 from the ACS. S.U. is a Rita Allen Foundation scholar and M.J.B. a trainee of the Medical Scientist Training Program (ST32 GM 7288). Part of the work is taken from a thesis to be submitted at Yeshiva University.

Received 23 December 1980; accepted 16 July 1981.

- Ginsberg, H. in *Virology*, 1149–1154 (Harper & Row, Maryland, 1980).
- Kilbourne, E. D. in *Textbook of Medicine* (eds Beeson, P. B., McDermott, W. & Wyngaarden, J. B.) 246–250 (W. B. Saunders, Pennsylvania, 1979).
- Panum, P. L. *Virchow's Arch. path. Anat. Physiol.* **1**, 492–512 (1847).
- McCarthy, K. *Br. med. Bull.* **15**, 201–204 (1959).
- Fraser, K. B. & Martin, S. J. in *Measles Virus and its Biology* (eds Tinsley, T. W. & Brown, F.) 25–40 (Academic, London, 1978).
- Graves, M. C., Silver, S. M. & Choppin, P. W. *Virology* **86**, 254–263 (1977).
- Birrer, M., Bloom, B. R. & Udem, S. *Virology* **108**, 381–390 (1981).
- Rowlands, D. *et al. Cell* **19**, 871–880 (1980).
- Laver, W. G., Gerhard, W., Webster, R. B., Frankel, M. E. & Air, G. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1425–1429 (1979).
- Wiktor, T. J. & Koprowski, H. *J. exp. Med.* **152**, 99–112 (1980).
- Fisher, L. E. & Rapp, F. J. *Viol.* **30**, 64–68 (1979).
- Hall, W. W., Lamb, R. A. & Choppin, P. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2047–2051 (1979).
- Horta-Barbosa, L., Fuccillo, D. A., Lonclon, W. T., Jabbove, J. T. J. & Sever, J. L. *Proc. Soc. exp. Biol. Med.* **132**, 272–277 (1969).
- Payne, F. E., Baublis, J. V. & Itabashi, H. H. *N. Engl. J. Med.* **281**, 585–590 (1969).
- Brown, J. L., Kato, K., Silver, J. & Nathenson, S. G. *Biochemistry* **13**, 3174 (1974).

Killing of *Mycobacterium microti* by immunologically activated macrophages

Linda Walker & D. B. Lowrie

MRC Unit for Laboratory Studies of Tuberculosis, Royal Postgraduate Medical School, Ducane Road, London W12 0HS, UK

Pathogenic mycobacteria such as tubercle bacilli are usually killed when immunity is developed in an infected animal or man. It has long been thought that macrophages are responsible for this killing but all attempts to demonstrate this *in vitro* have been unsuccessful^{1–7}, which has severely hampered our understanding of how immunity to tuberculosis is expressed. In particular, indirect evidence that hydrogen peroxide from macrophages might kill tubercle bacilli *in vivo*^{8–11} could not be properly tested. Recent studies have shown that macrophages can be activated by exposure to soluble products of immunologically stimulated T lymphocytes (lymphokines) so that they acquire a new or enhanced ability to kill infectious agents and tumour cells^{12–15}. Hydrogen peroxide has been implicated as a key macrophage product needed for killing^{13–16}. We report here that normal mouse peritoneal macrophages can be activated *in vitro* by lymphokines to kill *Mycobacterium microti*, the natural agent of tuberculosis in voles, and that added catalase protects the bacilli from killing.

We prepared lymphokine-rich supernatants of immunologically stimulated spleen cells (spleen cell factor, SCF) by a modification of a technique described elsewhere (see Fig. 1 legend and ref. 17). Macrophages that were maintained in the presence of SCF showed within 24 h morphological signs of activation in comparison to control, SCF-free, macrophages. The changes, which included symmetrical spreading and an increase in the number of phase-lucent vesicles and lysosomes, resembled those produced by macrophage-activating factor^{12,18}. The SCF-treated macrophages also released more hydrogen peroxide in comparison to untreated cells when stimulated with phorbol myristate acetate as described elsewhere^{13,14}.

In numerous experiments using five different batches of SCF, an enhanced killing of intracellular *M. microti* was found for SCF-activated macrophages compared with control macrophages. The percentage killed within 24 h of phagocytosis varied, depending on the batch of SCF used, with its concentration and the duration of macrophage exposure to SCF before phagocytosis. The best batches of SCF, when used in optimum concentrations (between 15 and 50%) for 3 days, activated macrophages to kill 70–97% of the intracellular bacilli within 24 h (Figs 1 and 2). In contrast, no more than bacteriostasis was usually observed in control macrophages. Decline in bacterial numbers in monolayers was not due to detachment of infected macrophages as assessed visually and by DNA estimation. Repeated rinsing of the monolayers kept the numbers of bacteria in monolayer supernatants below 10% of the numbers found in the monolayers themselves, thus there was no complication of interpretation imposed by continuing phagocytosis.

This induction of killing of *M. microti* in normal macrophages by treatment with SCF contrasts sharply with results from earlier attempts to evoke killing of pathogenic mycobacteria in mononuclear phagocytes by *in vitro* lymphokine treatment. Previous best results were an inhibition of growth of *Mycobacterium tuberculosis*^{1–5} or a transient period of slow killing of *Mycobacterium lepraemurium* in mouse cells¹⁹ or inhibition of *M. lepraemurium* or *M. microti* in human and rabbit cells⁷. Differences between this and previous studies are too numerous to permit identification of the key variable(s) nor can anything definite be said of the nature of the active constituent of SCF.

To assess whether hydrogen peroxide from the SCF-activated macrophages killed *M. microti*, we added either phorbol myristate acetate or catalase to the macrophages. Although phorbol

myristate acetate stimulates release of hydrogen peroxide from phagocytes in a manner analogous to a phagocytic stimulus²⁰, its presence during phagocytosis did not affect uptake of viable bacteria by the macrophages or the subsequent rate of bacterial death (Fig. 1). This type of soluble stimulus promotes phagocytic killing of other bacteria^{21,22}, but a negative result might only mean that hydrogen peroxide release into the bacterial micro-environment was maximally evoked by the stimulus of phagocytosis itself. Phagocytosis of an average of four *M. microti* bacilli per cell resembled stimulation with phorbol myristate acetate in evoking release of ~ 0.25 nmol $\text{H}_2\text{O}_2 \text{ min}^{-1}$ per 10^6 macrophages.

The fact that exogenous catalase protected *M. microti* from intracellular killing (Fig. 2) provides evidence for hydrogen peroxide being essential for killing. This agrees with findings for SCF-induced killing of protozoa within macrophages¹³. *M. microti* resembles fresh isolates of virulent *M. tuberculosis* in

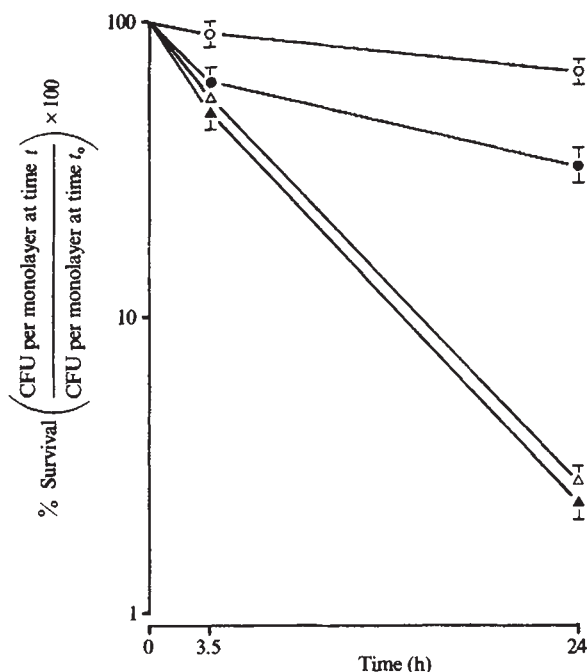


Fig. 1 Killing of *Mycobacterium microti* in SCF-activated macrophages. SCF was made using a published technique¹⁷ with the following minor modifications. Female CFLP mice (20–30 g) were injected intravenously with 3×10^7 viable *Mycobacterium bovis* BCG (Glaxo, UK) and after 4–5 weeks the spleens were removed and used to prepare a cell suspension in ice-cold medium 109 (Gibco-Biocult, UK). The cells were washed once in 109, resuspended (2×10^7 cells ml^{-1}) in 109 containing fetal calf serum (2% v/v), mercaptoethanol (5×10^{-3} M), L-glutamine ($100 \mu\text{g ml}^{-1}$) and purified protein derivative of tuberculin (PPD, $50 \mu\text{g ml}^{-1}$; Evans Medical, UK) then incubated in plastic Petri dishes at 37°C in 5% CO_2 atmosphere for 48 h. The supernatants (SCF) were then filtered through a $0.45\text{-}\mu\text{m}$ pore membrane and stored at -20°C . *M. microti* strain OV254 was grown in continuous culture in a chemostat (dilution rate $D = 0.028 \text{ h}^{-1}$) as previously described²⁴, except that the medium contained 0.2% (v/v) Tween 80 and iron, as ammonium ferric citrate, at a concentration of $2.0 \mu\text{g Fe}^{3+} \text{ ml}^{-1}$. Peritoneal macrophages were taken from female BALB/c mice (20–30 g) and kept for 3 days before infection as monolayers in plastic Petri dishes in medium alone²⁴ (○, ●), or a mixture of medium and 50% SCF (△, ▲). The medium or medium/SCF mixture contained benzylpenicillin at a concentration ineffective against *M. microti* (100 U ml^{-1}) and was changed daily. Monolayers were then overlaid with a *M. microti* suspension (2.5×10^7 bacteria ml^{-1}) in Hank's balanced salt solution (BSS) with or without phorbol myristate acetate ($1 \mu\text{g ml}^{-1}$; open and closed symbols respectively) and incubated at 37°C for 2 h to allow phagocytosis. The monolayers were then rinsed 4 times with warm BSS to remove extracellular bacteria and fresh medium or medium/SCF added to the dishes. Rinsing and medium replacement were repeated at intervals during incubation and at the same time the numbers of viable bacteria present in monolayers and supernatants were determined. Macrophages were scraped free and disrupted in digitonin (0.8 mg ml^{-1} ; nontoxic to bacteria) and bacterial colony-forming-units (CFU) counted on 7H11 agar⁸. Immediately after phagocytosis (time, $t = 0$) monolayers contained 2–10 bacteria per macrophage (macrophage number estimated from monolayer DNA content). Only the CFU counts in monolayers are shown, expressed as a percentage of those at time t_0 . Each point represents the geometric mean and range values (bars) from duplicate dishes.

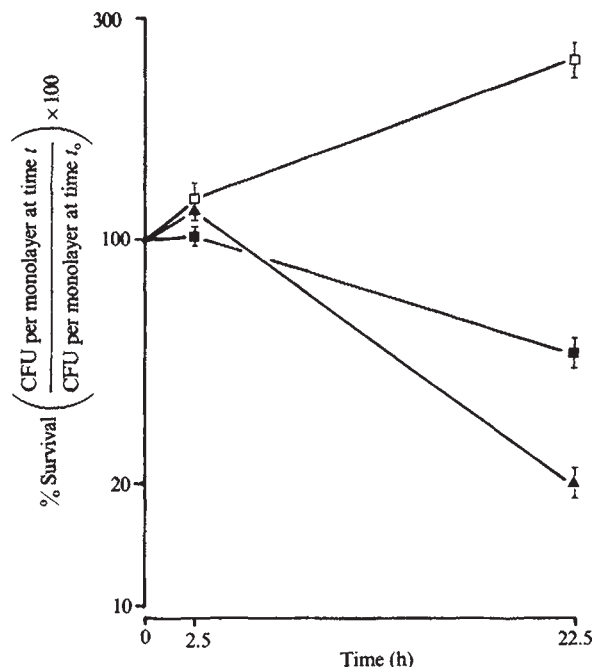


Fig. 2 The effect of catalase on killing of *M. microti* in SCF-activated macrophages. Before phagocytosis the macrophages were kept as monolayers for 24 h in medium containing 50% SCF; no antibiotics were present at any stage. *M. microti* was grown in a chemostat in medium containing $0.05 \mu\text{g Fe}^{3+} \text{ ml}^{-1}$. Catalase (2.5 mg ml^{-1} ; Sigma) was absent (△) or present in the monolayer dishes for 2 h before, during and after phagocytosis (□). Heat-denatured catalase (100°C for 30 min) was used in the same way (■). Immediately after phagocytosis, monolayers contained 1.3–3.6 bacteria per macrophage. Other experimental details were as described in Fig. 1 legend.

both catalase content ($\sim 11 \text{ mU}$ per mg extracted protein) and degree of susceptibility to H_2O_2 (ref. 8).

Host resistance to pathogenic mycobacteria undoubtedly involves multiple factors. For example, it is not obvious how hydrogen peroxide could cause the bacteriostasis which is a prominent feature of resistance in some forms of tuberculosis²³. However, the conclusion that macrophages can both kill *M. tuberculosis in vivo* and use hydrogen peroxide to do so, which was developed from circumstantial evidence, is now greatly strengthened.

We thank Professor D. A. Mitchison for helpful discussion and the MRC for financial support for L. W.

Received 16 February; accepted 29 June 1981.

- Patterson, R. J. & Youmans, G. P. *Infect. Immunity* **1**, 600–603 (1970).
- Klun, C. L. & Youmans, G. P. *J. Reticuloendothel. Soc.* **13**, 263–274, 275–285 (1973).
- Cahall, D. L. & Youmans, G. P. *Infect. Immunity* **12**, 851–857 (1975).
- Muroaka, S., Takeya, K. & Nomoto, K. *Jap. J. Microbiol.* **20**, 115–122, 365–373 (1976).
- Turcotte, R., DesOrmeaux, Y. & Borduas, A. F. *Infect. Immunity* **14**, 337–344 (1976).
- Alexander, J. & Smith, C. C. *Infect. Immunity* **22**, 631–636 (1978).
- Godal, T., Rees, R. J. W. & Lamvik, J. O. *Clin. exp. Immun.* **8**, 625–637 (1971).
- Jackett, P. S., Aber, V. R. & Lowrie, D. B. *J. gen. Microbiol.* **104**, 37–45 (1978).
- Jackett, P. S., Aber, V. R. & Lowrie, D. B. *Soc. gen. Microbiol.* **7**, 190 (1980).
- Jackett, P. S., Aber, V. R., Mitchison, D. A. & Lowrie, D. B. *Br. J. exp. Path.* **62**, 34–40 (1981).
- Jackett, P. S., Andrew, P. W., Aber, V. R. & Lowrie, D. B. *Br. J. exp. Path.* **62**, 419–428 (1981).
- David, J. R. & Remold, H. G. in *Biology of the Lymphokines* (eds Cohen, S., Pick, E. & Oppenheim, J. J.) (Academic, New York, 1979).
- Murray, H. W., Juangbhanich, C. W., Nathan, C. F. & Cohn, Z. A. *J. exp. Med.* **150**, 950–964 (1979).
- Nathan, C. F., Nogueira, N., Juangbhanich, C. W., Ellis, J. & Cohn, Z. A. *J. exp. Med.* **149**, 1056–1068 (1979).
- Buchmüller, Y. & Mauel, J. *Eur. J. Immun.* **10**, 428–433 (1980); *Experientia* **36**, 739 (1980).
- Nathan, C. F., Silverstein, S. C., Brukner, L. H. & Cohn, Z. A. *J. exp. Med.* **149**, 100–113 (1979).
- Nogueira, N. & Cohn, Z. A. *J. exp. Med.* **148**, 288–300 (1978).
- Nabarra, B., Dy, M., Andrianarison, I. & Dimitriu, A. *J. Reticuloendothel. Soc.* **25**, 73–83 (1979).
- Alexander, J. & Smith, C. C. *Infect. Immunity* **22**, 631–636 (1978).
- Nathan, C. F., Brukner, L. H., Silverstein, S. C. & Cohn, Z. A. *J. exp. Med.* **149**, 84–99 (1979).
- Issekutz, A. C., Lee, K.-Y. & Biggar, W. D. *Infect. Immunity* **24**, 295–301 (1979).
- Guarneri, J. J. & Goldstein, J. *Abstr. Am. Soc. Microbiol. 80th ann. Meeting*, 40 (1980).
- Hart, P. D. A. & Rees, R. J. W. *Br. J. exp. Path.* **41**, 414–421 (1960).
- Lowrie, D. B., Aber, V. R. & Carrol, M. E. *W. J. gen. Microbiol.* **110**, 409–419 (1979).

Immunoreactive substance P and serotonin present in the same dense-core vesicles

G. Pelletier, H. W. M. Steinbusch*
& A. A. J. Verhofstad*

MRC Group in Molecular Endocrinology, Le Centre Hospitalier de l'Université Laval, Quebec, Canada G1V 4G2

* Department of Anatomy and Embryology, University of Nijmegen, PO Box 9101, 6525EZ, Nijmegen, The Netherlands

It is well known that several endocrine cell types, particularly those found in the gastrointestinal tract, contain not only peptidic hormones but also a biogenic amine^{1,2}. Furthermore, it has recently been reported that a neuropeptide, substance P, and a classical neurotransmitter, serotonin, have been found in the same neurones in the central nervous system^{3,4}. However, these light microscopy studies were unable to elucidate the sites of storage of the two substances. Moreover, there is no information about the organelles involved in the storage of neuropeptides and biogenic amines which coexist in the same neurones, although it is well established that neuropeptides are mostly found in dense-core vesicles of terminals⁵⁻⁷. To confirm these previous results and to identify the storage organelles, we localized substance P and serotonin on consecutive ultrathin sections by immuno-electron microscopy. As we report here, we found that, in both raphe nuclei and dorsal horn of the spinal cord, no more than 20% of serotonin-containing nerve terminals were also positive for substance P. In these terminals, both substances were found in the same large dense-core vesicles (60–90 nm diameter). These results strongly suggest that, in appropriate conditions, the two substances are released simultaneously.

Adult male rats were fixed by perfusion in 300 ml of glutaraldehyde in 0.1 M cacodylate buffer (pH 7.6); raphe nuclei of the medulla oblongata and ventral horn of the spinal cord were carefully dissected and post-fixed by immersion for 1 h in 0.5% OsO₄ diluted in the same cacodylate buffer. Serial ultrathin sections from these regions were alternately stained for substance P and serotonin according to techniques described elsewhere^{8,9}. Antisera to substance P¹⁰ and serotonin¹¹ were used at a dilution in the range 1:2,000–1:4,000. The specificity of immunostaining was determined by absorption of each antiserum (diluted 1:2,000) with substance P and serotonin at a final concentration of 10⁻⁶ M. As previously reported^{10,11}, immunostaining was abolished only after absorption with the corresponding antigen.

We observed several terminals containing serotonin-immunoreactive material in the raphe nuclei and spinal cord. In these terminals, most of the reaction product was found in dense-core vesicles of diameter 60–90 nm. We could not detect any staining in small, clear vesicles. In each area examined ~40 serotonin-containing nerve ends were visualized; ~50% of these were seen making typical synaptic junctions. Adjacent sections were observed for staining for substance P; no more than 20% of these serotonin-containing terminals were positive for substance P in both raphe nuclei and spinal cord. The staining pattern was very similar for both antisera, all the dense-core vesicles being labelled (Fig. 1). Despite the difficulties in sectioning 60–90-nm dense-core vesicles, we were occasionally able to observe the same dense-core vesicles on two consecutive sections; staining for substance P was observed in vesicles that also contain serotonin.

This localization of serotonin clearly shows that serotonin immunoreactivity is predominantly contained in the dense-core vesicles of positive nerve endings. However, as certain amounts of immunoreactive material may have been lost or destroyed during the aqueous fixation procedures, these results do not

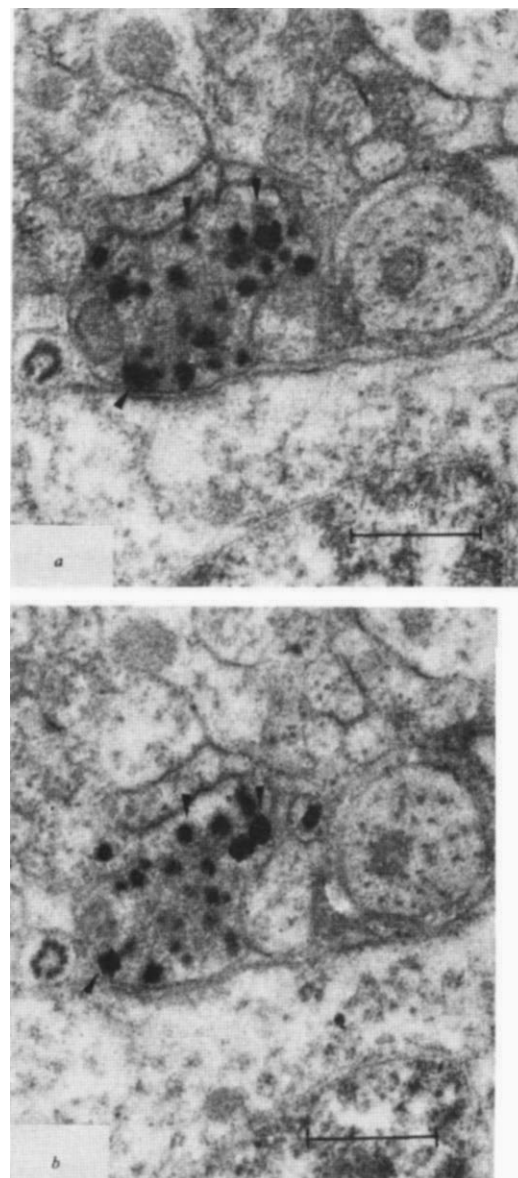


Fig. 1 *a*, Ultrathin section through the anterior horn of the spinal cord immunologically stained for serotonin. A strong reaction was found in all large dense-core vesicles (arrowheads) in a terminal whereas there was no detectable staining of small, clear vesicles. Note the absence of synaptic contact. *b*, Adjacent section stained for substance P. A strong reaction was also observed for all large dense-core vesicles (arrowheads) with no staining of small, empty vesicles. Scale bars, 0.4 μ m.

exclude the possibility that serotonin can be associated with other organelles such as small, clear vesicles.

Our results agree with previous observations^{3,4} that substance P and serotonin are localized in the same terminals. It is therefore very likely that these substances are simultaneously released when the terminal is stimulated. The concomitant release of peptides and biogenic amines has not been reported in the central nervous system. However, Viveros *et al.*¹² have shown that in the adrenal medulla, enkephalins and catecholamines are stored in the same secretory granules. It is possible that in the central nervous system some neurones could behave in the same way as cells of adrenal medulla, which are of neural origin.

The absence of typical synaptic junctions in many of the terminals which contain both substance P and serotonin suggests that they could be released at sites other than the synaptic junctions. This has already been proposed for the noradrenergic system¹³. If it is the case that a peptide and a biogenic amine are

released simultaneously in the brain, as suggested by our results, the physiological significance of such release remains to be discovered.

Received 14 April; accepted 3 July 1981.

- Pearse, A. G. E. *J. Histochem. Cytochem.* **17**, 303–313 (1969).
- Pearse, A. G. E. in *Chromaffin, Entero-chromaffin and Related Cells* (eds Coupland, R. E. & Fujita, T.) 147–154 (Elsevier, Amsterdam, 1976).
- Chan-Palay, V., Palay, S. L. & Jonsson, G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1582–1586 (1978).
- Hökfelt, T. et al. in *Neural Peptides and Neuronal Communications* (eds Costa, E. & Trabucchi, M.) 1–23 (Raven, New York, 1980).
- Pickel, V. M., Joh, T. H., Reis, D. J. & Leeman, J. E. *Brain Res.* **160**, 387–400 (1979).
- Pelletier, G. in *Central Nervous System* (eds Collu, R., Barbeau, A., Ducharme, J. & Rochefort, J. G.) 331–344 (Raven, New York, 1979).
- Larsson, L. I. *Histochemistry* **54**, 173–176 (1977).
- Sternberger, L. A., Hardy, P. H., Cuculies, J. J. & Mayer, H. G. *J. Histochem. Cytochem.* **18**, 315–333 (1970).
- Moriarty, G. C. & Halmi, N. S. *J. Histochem. Cytochem.* **20**, 590–623 (1972).
- Pelletier, G., Leclerc, R. & Dupont, A. *J. Histochem. Cytochem.* **25**, 1373–1375 (1977).
- Steinbusch, H. W. M. & Verhofstad, A. A. J. in *Advances in Pharmacology and Therapeutics* Vol. 2 (ed. Simon, P.) 151–160 (Pergamon, Oxford and New York, 1978).
- Viveros, O. H., Diliberto, E. J., Hazum, E. & Chang, K. J. in *Neural Peptides and Neuronal Communications* (eds Costa, E. & Trabucchi, M.) 191–204 (Raven, New York, 1980).
- Descarries, L., Watkins, K. C. & Lapierre, Y. *Brain Res.* **133**, 197–222 (1977).

Mouse skin cells resistant to terminal differentiation associated with initiation of carcinogenesis

Stuart H. Yuspa & David L. Morgan

Laboratory of Experimental Pathology, National Cancer Institute, Bethesda, Maryland 20205, USA

In chemical carcinogenesis, initiation is defined as the exposure of a tissue, generally mouse skin, to a limited dose of carcinogen which by itself does not induce tumours¹. Subsequent treatments, usually by tumour promoters, are required to allow expression of initiation. In the absence of further treatment, the initiated cells are not recognizable by any known biochemical or biological criteria and the tissue itself functions normally. Much research has centred on elucidating the mechanism of initiation^{1–3}. It seems likely that initiation involves a genetic change in the target cell as its expression after tumour promotion is a permanent characteristic of the tissue, even if many cell generations have elapsed between initiation and promotion⁴. Studies on tissues from patients with familial polyposis coli suggest that the occurrence of initiated cells in these individuals is due to an autosomal dominant gene⁵. Few studies have actually attempted to define the biological nature of the initiating alteration itself. We report here that cells resistant to terminal differentiation can be readily isolated from skin of BALB/c mice exposed to an initiating dose of carcinogen *in vivo* but not from control mouse skin. Also, we found that one strain of mouse (SENCAR) might contain a preexisting population of cells resistant to differentiation, and possibly initiated for tumorigenesis.

The association of initiation with an altered response to differentiation signals was suggested by our observations that exposure of cultured mouse epidermal basal cells to initiators of carcinogenesis yields cellular foci which resist the proliferative block associated with induced terminal differentiation. Carcinogen-altered cells proliferated in culture conditions where normal cells differentiated terminally and died^{6–8}. Cells from such foci were not tumorigenic, but malignant epidermal cells also resisted the differentiation signal *in vitro*^{6,9}. Our studies took advantage of the discovery that extracellular calcium concentration regulates epidermal growth and differentiation in cell culture¹⁰—in medium with 0.02–0.09 mM Ca²⁺, epidermal cells maintain a high proliferation rate and have characteristics of basal cells¹¹, but when medium Ca²⁺ is increased (1.2–1.4 mM), proliferation stops and terminal differentiation proceeds through stratification and cornification, leading to cell death and sloughing from the culture dish. Altered cells can be easily selected because they proliferate in conditions (1.2 mM Ca²⁺) where normal cells have sloughed. After induced differentiation, dishes can be fixed, stained and positive plates or foci counted⁶.

Table 1 Differentiation-resistant (rhodamine-positive) colonies from mouse epidermis treated *in vivo* with initiating dose of 7,12-dimethylbenz[a]anthracene (DMBA)

Experiment	<i>In vivo</i> treatment	Time <i>in vitro</i> (weeks)		Positive plates/total plates
		0.02 mM Ca ²⁺	1.4 mM Ca ²⁺	
308RB (BALB/c)	TPA (5 µg twice weekly × 4 weeks)	11	4	0/10
	Acetone (200 µl twice weekly × 4 weeks)	11	4	1/23*
308RB1 (BALB/c)	DMBA (25 µg × 1; no treatment × 2 weeks)	11	4	4/11†
	Acetone (200 µl × 1; no treatment × 2 weeks)	11	4	0/9
308RB2 (BALB/c)	DMBA (25 µg × 1)	11	4	14/31†
	Acetone (200 µl twice weekly × 4 weeks)	11	4	24/32†
	TPA (5 µg twice weekly × 4 weeks)	11	4	0/30
	Acetone (200 µl × 1)	11	4	1/27*
	Acetone (200 µl twice weekly × 4 weeks)	11	4	18/30†
	DMBA (25 µg × 1)	11	4	7/29*†
308RS (SENCAR)	TPA (5 µg twice weekly × 4 weeks)	11	4	7/29*
	Acetone (200 µl × 1)	11	4	7/29*
	TPA (5 µg twice weekly × 4 weeks)	11	4	7/29*
	Acetone (200 µl twice weekly × 4 weeks)	11	4	7/29*

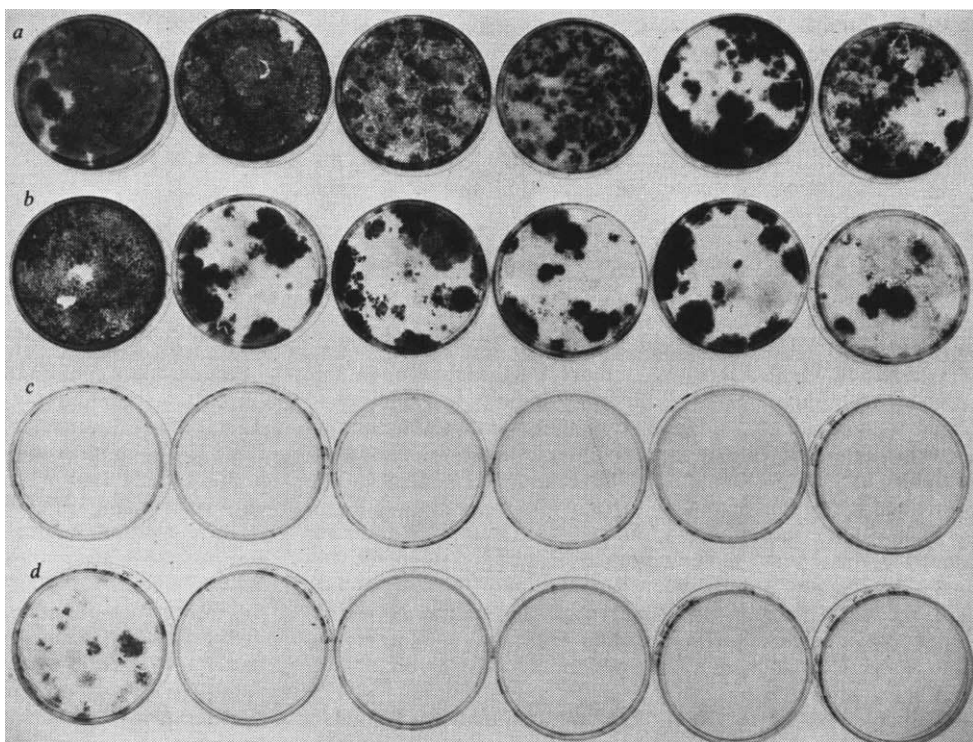
Epidermal cells were prepared from skins of four to six animals per group treated *in vivo* as shown in the second column. The methods for isolation of keratinocytes from adult mice have been described previously⁷. Pooled cell preparations were counted in a haemocytometer with Trypan blue and 2 × 10⁶ viable cells plated per 35-mm culture dish in low calcium Eagle's medium (LCEM) (Eagle's minimal essential medium prepared without calcium and containing non-essential amino acids (× 1) (Microbiological Associates) and 8% Chelex-treated fetal calf serum (FCS, Reheis) and 1% antimycotic-antibiotic (Gibco)¹⁰. After 24 h, cultures were washed and re-fed with LCEM and maintained in this medium (which has a Ca²⁺ content of 0.02 mM and supports the growth of basal cells) for 10 weeks with medium changes twice weekly. At this time cultures from all groups were generally 50–100% confluent. Cultures were then switched to medium 199, containing 2% FCS and 0.07 mM Ca²⁺ for 1 week and then medium 199, containing 2% FCS and 1.2 mM Ca²⁺ for 1 week. Finally, all cultures were switched to Eagle's medium containing 8% FCS and 1.4 mM Ca²⁺. These medium changes enhance induced terminal differentiation in normal cells and have been discussed previously⁶. After 4 weeks in 1.2–1.4 mM Ca²⁺ medium, plates were fixed with 10% formalin and stained with rhodamine, which stains stratifying epidermal colonies red⁶. Plates were scored as positive or negative depending on whether there were red colonies > 5 mm in diameter. All colonies were also examined using a dissecting microscope to ensure that red foci were not residual keratinized debris.

*Small colonies (< 1 cm).

†Large colonies (> 1 cm) or confluent dish.

The dorsal epidermis of BALB/c or SENCAR mice, 7–8 weeks old and in the resting phase of the hair cycle, was painted once with 25 µg of 7, 12-dimethylbenz[a]anthracene (DMBA) in 200 µl of acetone. Control mice received acetone only. This amount of DMBA is insufficient to produce tumours in the absence of promotion, but if promotion follows, many tumours develop with a T₅₀ (50% tumour incidence) latent period of 8–10 weeks for SENCAR and 20–30 weeks for BALB/c. Thus, we define 25 µg DMBA as an initiating dose of carcinogen for mouse skin. In some animals, the tumour promoter 12-O-tetradecanoyl phorbol-13-acetate (TPA), 5 µg in 200 µl acetone, was applied twice weekly beginning 7 days after initiation. Controls received acetone only. TPA or acetone exposures were continued for 4 weeks, not long enough for tumours to develop. Either 3 days or 1 week after the last application of acetone or TPA (6 weeks after DMBA), animals were killed and

Fig. 1 Rhodamine-stained plates representative of experiment 308RB2. *In vivo* exposure to: *a*, both DMBA and TPA; *b*, DMBA only; *c*, TPA only; *d*, acetone only. Positive plate in *d* contains the most definitive colonies seen in a BALB/c control.



the dorsal skin removed for culture. In one experiment (308 RB1, Table 1) animals were killed 2 weeks after DMBA without TPA treatment. Culture conditions which allow selection of differentiation-resistant foci are described in Table 1 legend.

Table 1 gives the results of four experiments. Only 2 out of 99 dishes derived from epidermis of non-initiated BALB/c mice with or without TPA contained foci resistant to Ca^{2+} -induced terminal differentiation and both had only very small, relatively pale colonies (Fig. 1*d*). In the initiated BALB/c groups, 50–60% of all dishes contained colonies resistant to terminal differentiation and most were either confluent or contained multiple large, dark red colonies (Fig. 1*a, b*). The intensity of the red staining reflects both the cell number in each colony and the degree of vertical stratification. Because of the difficulty in distinguishing individual colonies, we did not attempt to quan-

tify them. Experiment 308 RB1 shows that resistant colonies can be expressed in culture within 2 weeks of *in vivo* initiation. This finding does not exclude the possibility that changes begun *in vivo* could have been completed during the 11 weeks of *in vitro* growth in 0.02 mM Ca^{2+} .

Table 1 also shows the results of a similar protocol with SENCAR mice. These mice have been bred for sensitivity to skin carcinogenesis and we have previously suggested that initiated cells may pre-exist in their epidermis^{7,12}. In this experiment the initiated group (DMBA only) yielded ~60% positive plates with many large red colonies (Fig. 2*b*). Cultures from SENCAR animals receiving both DMBA and TPA had a smaller number of positive plates but in general the colonies resembled those of the group receiving DMBA alone (three confluent plates from this group are not shown because they were subcultured to

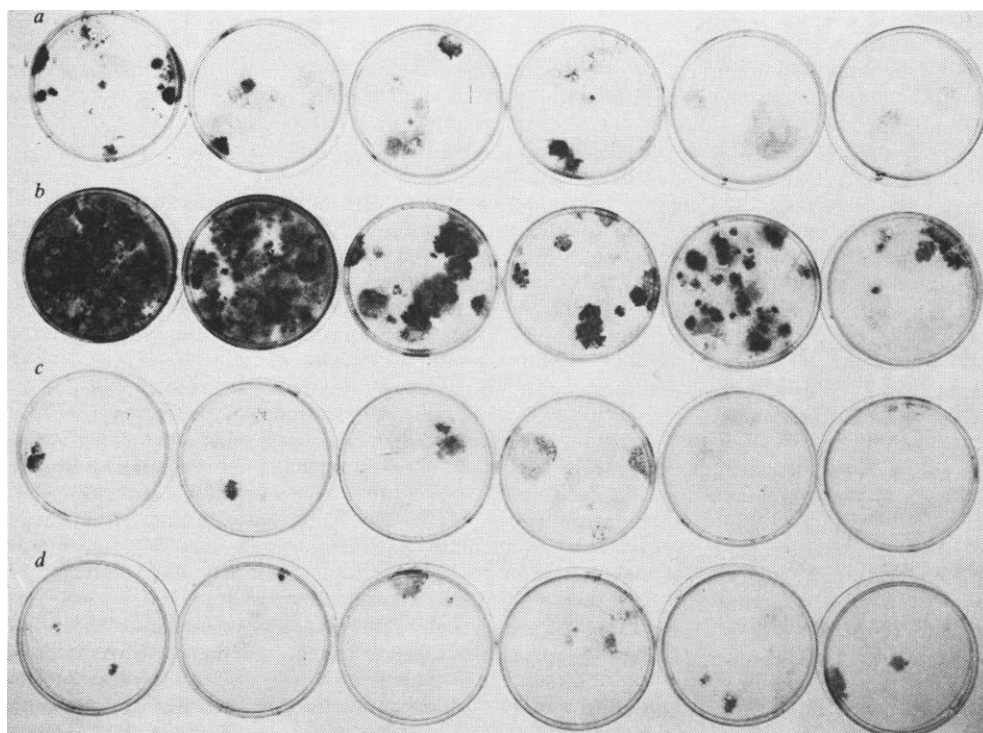


Fig. 2 Rhodamine-stained plates representative of experiment 308RS. *In vivo* exposure as for Fig. 1. Not shown in *a* are three confluent plates used for isolating colonies for further study. Two plates on far right of *a* and *c* are negative. Small colonies in *c* and *d* were also evaluated by microscopic appearance to establish that they were positive.

obtain cell lines). This reduction in positive plates could be due to the skin cytotoxicity associated with exposure of SENCAR mice (but not BALB/c mice) to 5 μ g TPA¹². Experiment 308 RS indicates that uninitiated control groups yielded positive plates (Table 1). Although the colonies in control plates were qualitatively different from those in DMBA-initiated groups (Fig. 2c, d), microscopic analysis indicated that they were surviving and growing colonies. A second experiment with SENCAR cells (not shown here) with a slightly varied protocol confirmed that control plates were positive in the SENCAR group. Thus, untreated SENCAR mouse skin contains cells similar to initiated skin with regard to resistance to terminal differentiation. These data agree with our finding that TPA treatment of SENCAR mice (but not BALB/c mice) in the absence of initiation leads to the development of a significant number of papillomas¹². However, either a smaller number of cells have this trait or it is expressed to a lesser or qualitatively different extent in untreated SENCAR and can be significantly amplified by treatment with initiator.

In all studies, the colonies which persisted after Ca²⁺-induced differentiation had an epidermal morphology and could be subcultured. These cells seemed to be identical to terminal differentiation-resistant epidermal cells isolated from the *in vitro* assay previously reported⁶. Studies are now in progress to determine if the cells are tumorigenic and to characterize their biological properties.

Our studies indicate that initiation of carcinogenesis in mouse skin is associated with the evolution of cells which resist induced terminal differentiation by Ca²⁺ *in vitro*. *In vivo*, such cells would proliferate even when no longer in contact with the basement membrane, a condition which results in a proliferative block and terminal differentiation in normal cells. We have suggested a mechanism by which cells resistant to differentiation signals would clonally expand in response to tumour promoters that cause a regenerative hyperplasia^{8,13}. Such clones would be expected to have a high labelling index but still express differentiative function in the upper layers. These characteristics are found in papillomas, which are the principal product of two-stage carcinogenesis in skin.

Previous data¹⁴⁻¹⁷ suggest that an alteration in regulation of normal differentiation, particularly with regard to the proliferative block commonly associated with terminal differentiation, is an early event in chemical carcinogenesis, possibly the initiating event. The assay we have developed can recognize and quantitate this change, and also provides a way of isolating the altered population.

We thank Drs Henry Hennings and Umberto Saffiotti for helpful suggestions and Mrs Margaret Green for secretarial assistance.

Received 19 March; accepted 30 June 1981.

- Yuspa, S. H. & Harris, C. C. in *Cancer Epidemiology and Prevention* (eds Schottenfeld, D. & Fraumeni, J.) (Saunders, Philadelphia, in the press).
- Miller, E. C. *Cancer Res.* **38**, 1479-1496 (1978).
- Yuspa, S. H., Hennings, H. & Saffiotti, U. *J. invest. Derm.* **67**, 199-208 (1976).
- Van Duuren, B. L., Sivak, A., Katz, C., Seidman, I. & Melchionne, S. *Cancer Res.* **35**, 502-505 (1975).
- Kopelovitch, L. in *Colorectal Cancer; Prevention, Etiology and Screening* (eds Winawer, S., Schottenfeld, D. & Sherlock, P.) 97-108 (Raven, New York, 1981).
- Kulesz-Martin, M., Koehler, B., Hennings, H. & Yuspa, S. H. *Carcinogenesis* **1**, 995-1006 (1980).
- Yuspa, S. H., Licht, U., Morgan, D. & Hennings, H. in *Biochemistry of Normal and Abnormal Epidermal Differentiation* (eds Bernstein, I. A. & Seiji, M.) 171-188 (University of Tokyo Press, 1980).
- Yuspa, S. H., Hennings, H., Kulesz-Martin, M. & Licht, U. in *Cocarcinogenesis and Biological Effects of Tumour Promoters* (ed. Hecker, E.) (Raven, New York, in the press).
- Yuspa, S. H., Hawley-Nelson, P., Koehler, B. & Stanley, J. R. *Cancer Res.* **40**, 4694-4703 (1980).
- Hennings, H. *et al. Cell* **19**, 245-254 (1980).
- Yuspa, S. H., Hawley-Nelson, P., Stanley, J. R. & Hennings, H. *Transplant Proc.* **12**, Suppl. 1, 114-122 (1980).
- Hennings, H. *et al. Cancer Res.* **41**, 773-779 (1981).
- Yuspa, S. H., Ben, T. B., Hennings, H. & Licht, U. *Biochem. biophys. Res. Commun.* **97**, 700-708 (1980).
- Thurnherr, N., Deschner, E., Stonehill, E. & Lipkin, M. *Cancer Res.* **33**, 940-945 (1973).
- Klein-Szanto, A. J. P., Nettesheim, P., Topping, D. C. & Olson, A. C. *Carcinogenesis* **1**, 1007-1016 (1980).
- Morson, B. & Bussey, H. *Curr. Probl. Surg.* **2**, 1-52 (1970).
- Kopelovitch, L., Bias, N. E. & Helson, L. *Nature* **282**, 619-621 (1979).

Metastatic potential correlates with cell-surface protein alterations in B16 melanoma variants

Manuel Rieber & Mary S. Rieber

Centre of Microbiology & Cell Biology,
Instituto Venezolano de Investigaciones Científicas,
Apartado 1827, Caracas 1010A, Venezuela

The invasive properties of metastatic cells may be related to their ability to form cell-cell and cell-endothelial matrix interactions, behaviour which may be mediated by tumour cell-surface proteins. However, efforts to find qualitative differences in surface protein between cells with different metastatic potential have been disappointing¹⁻⁵. The *in vivo* metastatic ability of B16 melanoma cell lines has been correlated with their ability to aggregate with platelets^{5,6} and 1 M urea extracts of metastatic cells are known to be more effective in platelet aggregation than those from their less metastatic counterparts⁷. In addition, recent *in vivo* experiments have shown that plasma membrane vesicles from highly metastatic tumour cells can modify the metastatic behaviour of poorly metastatic sublines when fused with the latter⁸. We have now examined possible differences in the surface macromolecules of B16 metastatic variants; our results show novel qualitative differences in surface protein that correlate with differing metastatic behaviour of the melanoma cells.

The experiments described here were done using B16-F10^{Lr} cells (low incidence of metastasis in the lung, selected *in vitro* for resistance to lysis by syngeneic lymphocytes)⁹, B16-F10 cells (high incidence of lung colonization, selected by successive passaging of lung tumour colonies for 10 *in vivo/in vitro* selections)¹⁰ and B16-BL6 selected from the B16-F10 cell line⁴ for increased invasive capacity. All the cell variants were maintained *in vitro* and assayed for metastasis *in vivo*, as described elsewhere^{5,11}. In our experiments we observed that intramuscular (i.m.) injection into 9-12-week old C57BL mice with 2.5×10^4 cells from subconfluent cultures produced multiple experimental pulmonary metastasis in 60-75% of the animals for the BL6 cell line, or clones derived from it. Compare this with 25-33% multiple lung metastases in animals exposed to F10 cells or their clones, and <5% pulmonary metastasis when animals were injected with cloned cultures derived from F10^{Lr} cells. These results agree with those reported elsewhere^{4,5}.

Radioiodination of surface proteins was carried out on cell monolayers using lactoperoxidase, glucose oxidase and ¹²⁵I-labelled NaI (1 mCi per 10⁷ cells), followed by solubilization, electrophoretic analysis in 7.5% SDS-gels and autoradiography, as described elsewhere¹². Enzymatic iodination of B16 melanoma cell-surface proteins and subsequent electrophoretic analysis showed labelling mostly in the 78,000 (78 K), 140 K, 200 K and 45 K molecular weight (MW) regions but failed to reveal significant qualitative differences between cells of differing metastatic potential (Fig. 1, lanes a-c). Nevertheless, some minor quantitative differences were observed in the B16-F10^{Lr} cells, used as a control because of their markedly lower metastatic potential (lane c). These findings agree with previous reports on the same cell lines⁵. Similar results were obtained using sodium periodate-tritiated borohydride labelling (data not shown); this also agrees with other reports⁵.

As 1 M urea is known to release platelet-aggregating ability from tumour cells^{6,7}, we examined whether such treatment of pre-iodinated cells led to a differential solubilization in cells of differing metastatic potential. A comparison of the molecules solubilized by 1 M urea extraction revealed that F10^{Lr} cells of

low metastatic potential gave a component of MW ~18 K, as well as some components in the 78 K-MW region. In contrast, components of MW ~78 K, and no low molecular weight species, were detected in the 1 M urea fraction of F10 metastatic cells (Fig. 1, lanes *d, e*).

As *p*-formaldehyde and dithiotreitol (DTT) have been shown to potentiate plasma membrane vesiculation^{13,14} and because plasma membrane vesicles from metastatic but not from non-metastatic cells were shown to be important in the metastatic process⁸, we examined the effect of conditions that promote membrane vesiculation on surface proteins from preiodinated metastatic BL6 and F10 cells (Fig. 1, lanes *f, g*) and on non-metastatic FLr cells (Fig. 1, lane *h*). The patterns shown by two metastatic cells were very similar (Fig. 1*f, g*) and different from that seen in the less metastatic variant (Fig. 1*h*). In the latter, we observed the presence of an 18 K-MW species, the absence of the 45 K-MW components seen in metastatic cells (Fig. 1*f, g*), and a significant radioactive species which migrated slightly faster than the 78 K-MW components of metastatic cells (more evident in Fig. 1, lane *i*, in which lower levels of radioactivity were used than in lane *h*).

To study the effect of urea on surface proteins, we compared the protein species solubilized after treatment of pre-iodinated cells with 1 M or 8 M urea. Contrary to the expectation that higher levels of urea would allow the isolation of greater quantities of 18 K-MW surface component from low metastatic F10^{lr} cells, 8 M urea extraction of pre-iodinated F10^{lr} cells gave lower levels of the fast migrating protein, whereas it was clearly detectable in 1 M urea extracts of iodinated cells (Fig. 2*b*). The fact that the 8 M urea extracts showed the presence of slower-migrating components (Fig. 2*a*) and did not reveal several faster migrating components evident in iodinated cells extracted in 1 M urea may be due to the release or activation by the lower urea concentration of a protease from non-metastatic cells; this protease may be inhibited in 8 M urea.

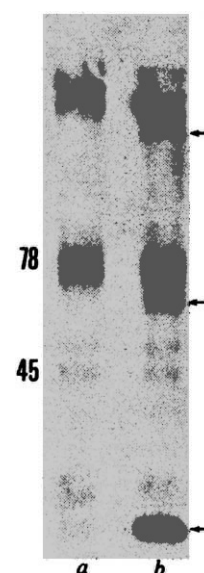


Fig. 2 Effect of urea on the surface-associated components of poorly metastatic cells. Iodinated F10^{lr} cells were extracted for 16 h at 4 °C with Tris-HCl (0.05 M, pH 7.4) containing either urea 8 M (*a*) or 1 M (*b*). After centrifugation of the extracts at top speed for 15 min in an Eppendorf microcentrifuge, the supernatants were subjected to electrophoretic analysis. The more obvious differences between *a* and *b* are indicated by arrows.

As tumour cell heterogeneity is known to play an important part in metastasis^{5,11}, we cloned the F10, BL6 and F10^{lr} cells at limiting dilution to test whether some of the reported differences could also be detected in their clones. A typical experiment, using conditions that promote membrane vesiculation^{13,14} in the newly derived sublines which retained the biological behaviour of their parent cells, once again showed increased migration in the 78 K-MW region (Fig. 3) as well as the other fast-migrating species present only in the less metastatic variants (Fig. 3, lanes *a-c*).

Cell monolayers were extracted with non-ionic detergents to prepare a cytoskeletal matrix, as we thought that the cytoskeleton might significantly affect surface behaviour¹⁵ and that such extraction might elucidate different solubilization properties of surface molecules for variants of differing metastatic behaviour. To test the latter possibility, we examined the effect of extracting surface-iodinated cells with isotonic medium containing Triton X-100 detergent by comparative electrophoretic analysis¹⁶. Whereas the Triton-insoluble components from iodinated metastatic BL6 and F10 cells gave essentially similar patterns with comparable proportions of species in the 78 K-MW region (Fig. 3*d, e*), the corresponding fraction from the less metastatic F10^{lr} variant showed a greater level of radioactivity in faster-migrating bands (Fig. 3*f*) and in particular in components with electrophoretic migration similar to the species that were extracted using 1 M urea from iodinated monolayers of the same low metastatic variant (see Fig. 2).

To our knowledge, this is the first report to show differences in cell-surface proteins between cells that produce significant experimental metastasis in the lung, and cells of low metastatic ability. These differences, demonstrated in conditions that promote membrane vesiculation, or after urea or Triton X-100 extraction of surface-iodinated cells, seem to be compatible with recent *in vivo* experiments⁸ which indicate that plasma membrane vesicles from metastatic cells differ in their *in vivo* behaviour from those of less metastatic variants.

It is also interesting that 1 M urea extraction of surface components revealed significant differences between the low and high metastatic variants. As surface components of metastatic cells extracted in this way have a role in platelet aggregation⁷, it is tempting to suggest that one reason the low

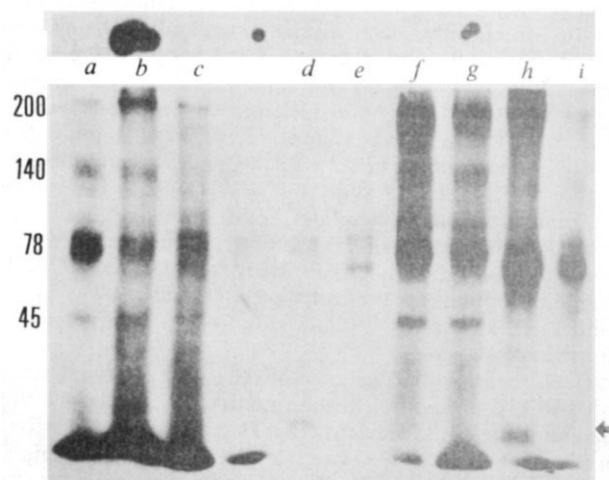


Fig. 1 Expression of cell-surface alterations in B16 melanoma variants. Cell monolayers were iodinated¹² and examined as such, or after exposure to *p*-formaldehyde/DTT^{13,14}, by solubilization and electrophoresis as described elsewhere¹². For extraction with 1 M urea, a 5X mixture was added to the iodinated cells to give a final concentration of 2% SDS, 0.1 M β -mercaptoethanol and 0.1 M Tris-HCl, pH 6.8, at 90 °C for 3 min before electrophoresis in 7.5% polyacrylamide gels¹². *a*, B16-BL6 cells; *b*, B16-F10 cells; *c*, B16-F10^{lr} cells; *d*, aliquot (~25%) of a 1 M urea extract of iodinated F10^{lr} cells; *e*, aliquot (~25%) of a 1 M urea extract of iodinated F10 cells; *f-h*, same cells as for *a-c* but treated with *p*-formaldehyde/DTT; *i*, same as in *h* but with lower levels of radioactivity. An arrow indicates the position of the 18 K-MW component seen in samples from cells of low metastatic potential.

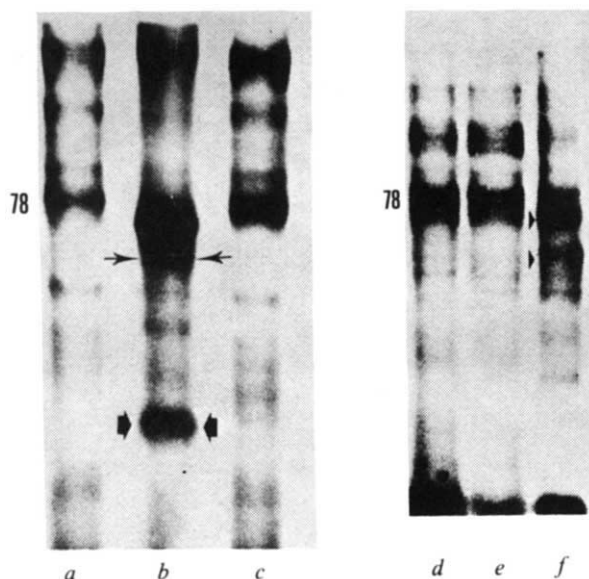


Fig. 3 Effect of treatment with *p*-formaldehyde/DTT and Triton X-100 on the surface proteins of cloned lines of B16 melanoma variants. Clones of B16 melanoma variants obtained by limiting dilution of the parent cells were cultured in microtitre plates to give stocks for surface iodination, exposed to *p*-formaldehyde/DTT (*a*–*c*) or Triton X-100 (*d*–*f*), and products separated by electrophoresis as described in the text. *a* and *d*, B16–BL6 clone; *b* and *e*, B16–F10^{Lr} clone; *c* and *f*, B16–F10 clone. The arrows show the positions of the more obvious changes in the non-metastatic clones. Similar results were obtained when five other clones of each cell type were tested for cell surface properties and experimental metastatic behaviour, which resembled that of the parent lines.

metastatic variants fail to promote the platelet aggregation that is important for spread of malignant disease^{6,7} is the preferential degradation of such components in low, compared with high, metastatic variants. This greater susceptibility of surface proteins from low metastatic cells was shown not only in conditions of induced membrane vesiculation or urea extraction, but also on extraction of labelled cells with Triton X-100 in conditions used to prepare cytoskeletal matrices¹⁶. When all three extraction procedures were used on poorly metastatic cells, we obtained components of lower molecular weight than those observed in the more metastatic cells.

An explanation consistent with these results is that cells with lower metastatic potential have greater protease and/or glycosidase activity, which in turn leads to a greater turnover of the surface components required for cell–cell interaction and subsequent cellular survival during the invasive process.

We thank Dr I. J. Fidler for providing the B16 melanoma cells used here, through the NCI Tumor Bank.

Evidence for association of glycosphingolipid with a colchicine-sensitive microtubule-like cytoskeletal structure of cultured cells

Kooko Sakakibara, Takashi Momoi, Tsutomu Uchida & Yoshitaka Nagai

Department of Pathobiochemical Cell Research, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108, Japan

Glycosphingolipids are generally considered to be present only on the surface of plasma membranes and to participate in surface-related cellular activity¹. However, our previous immunohistochemical studies² showed that affinity-purified antibody against galactocerebroside (GC) labelled not only the plasma membrane, but also the cytoplasm of epithelial cells in particular areas of hamster kidney, liver and lung. To confirm these findings and to try to obtain more detailed information on the subcellular localization of GC, we reinvestigated this problem in an epithelial cell line derived from monkey kidney. We report here the results indicating that GC is present in the cytoplasm of some cultured cell lines of epithelial origin in close association with colchicine-sensitive microtubule-like subcellular structure.

JTC-12, an established cell line derived from cynomolgus monkey kidney³, is known to be parathyroid hormone sensitive⁴ and to have γ -glutamyltranspeptidase activity⁵, suggesting its renal cortical epithelial origin. The cells were serially cultivated in DM-160 medium with 10% calf serum. Rabbit antibody against GC of bovine spinal cord was purified by affinity chromatography as described previously⁶. The antibody showed strong affinity for GC, and high specificity in that it did not react with any structurally related glycosphingolipids⁶. When assayed by complement fixation, the titre of the antibody preparation was 1:200. For indirect immunoperoxidase or immunofluorescent studies using anti-GC antibody, the cells were dispersed with 0.25% trypsin, plated on Lab-Tek chamber slide no. 4804 at a concentration of $2\text{--}5 \times 10^3$ cells per well and allowed to become attached to the slide for 24–48 h. Cells were washed with phosphate-buffered saline (PBS), immediately air dried (unfixed) or fixed for 10 min in acetone at room temperature and then incubated with anti-GC antibody for 30 min at 37°C. The antibodies bound to the cells were visualized by indirect immunoperoxidase or immunofluorescence methods as described elsewhere⁷.

JTC-12 cells, either fixed or unfixed, were positively stained with anti-GC antibody, showing numerous cytoplasmic fibres originating near the nucleus and extending out towards the periphery of the cell (Fig. 1*a*). Individual fibres were thin and separated from each other, following various courses; some were straight and some bent or curved. These patterns and the distribution of cytoplasmic fibres located with anti-GC antibody seemed to be very similar to those reported for microtubules or intermediate filaments^{8,9}. However, unlike microtubules, cells in the mitotic phase showed strong staining all over the cytoplasm, but not in mitotic spindles and intercellular bridge (Fig. 1*b*). When antibody preabsorbed with GC or non-immunized rabbit serum was substituted as control for anti-GC antibody, no fibrous structure were demonstrated in the cytoplasm (Fig. 1*c*).

To identify the cytoplasmic fibres stained with anti-GC antibody, cytoskeleton was prepared by the method of Osborn and Weber¹⁰, from cells in stabilization buffer containing 0.5% Triton-X 100, and then subjected to immunostaining for GC. As Fig. 2*a* shows, the cytoplasmic fibrous networks were most clearly visualized, indicating that GC is associated with cyto-

Received 18 February; accepted 18 June 1981.

1. Tao, T. W. & Burger, M. M. *Nature* **270**, 437–438 (1977).
2. Yogeewaran, G., Stein, S. B. & Sebastian, H. *Cancer Res.* **38**, 1336–1344 (1978).
3. Nicolson, G. L. *et al.* in *Cell and Tissue Interaction* (eds Lash, J. W. & Burger, M. M.) 225–241 (Raven, New York, 1977).
4. Hart, I. R. *Am. J. Path.* **97**, 587–600 (1979).
5. Raz, A. *et al.* *Cancer Res.* **40**, 1645–1651 (1980).
6. Gasic, G. J., Gasic, R. B., Galanti, N., Johnson, T. & Murphy, S. *Int. J. Cancer* **11**, 704–718 (1973).
7. Pearlstein, E., Salk, P. L., Yogeewaran, G. & Karparkin, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4336–4339 (1980).
8. Poste, G. & Nicholson, G. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 399–403 (1980).
9. Fidler, I. J., Gersten, D. M. & Budmer, M. B. *Cancer Res.* **36**, 3160–3165 (1976).
10. Fidler, I. J. *Cancer Res.* **35**, 218–224 (1975).
11. Raz, A., Bucana, C., McLellan, W. & Fidler, I. J. *Nature* **284**, 363–364 (1980).
12. Rieber, M. & Rieber, M. S. *Cancer Res.* **40**, 2562–2567 (1980).
13. Scott, R. E., Perkins, R. G., Zschunke, M. A., Hoerl, B. J. & Maercklin, P. E. *J. Cell Sci.* **35**, 229–243 (1979).
14. Scott, R. E. & Maercklin, P. B. *J. Cell Sci.* **35**, 245–252 (1979).
15. Nicolson, G. L. *Biochem. biophys. Acta* **450**, 1–72 (1976).
16. Brown, S., Levinson, W. & Spudich, J. S. *J. supramolec. Struct.* **5**, 119–130 (1976).

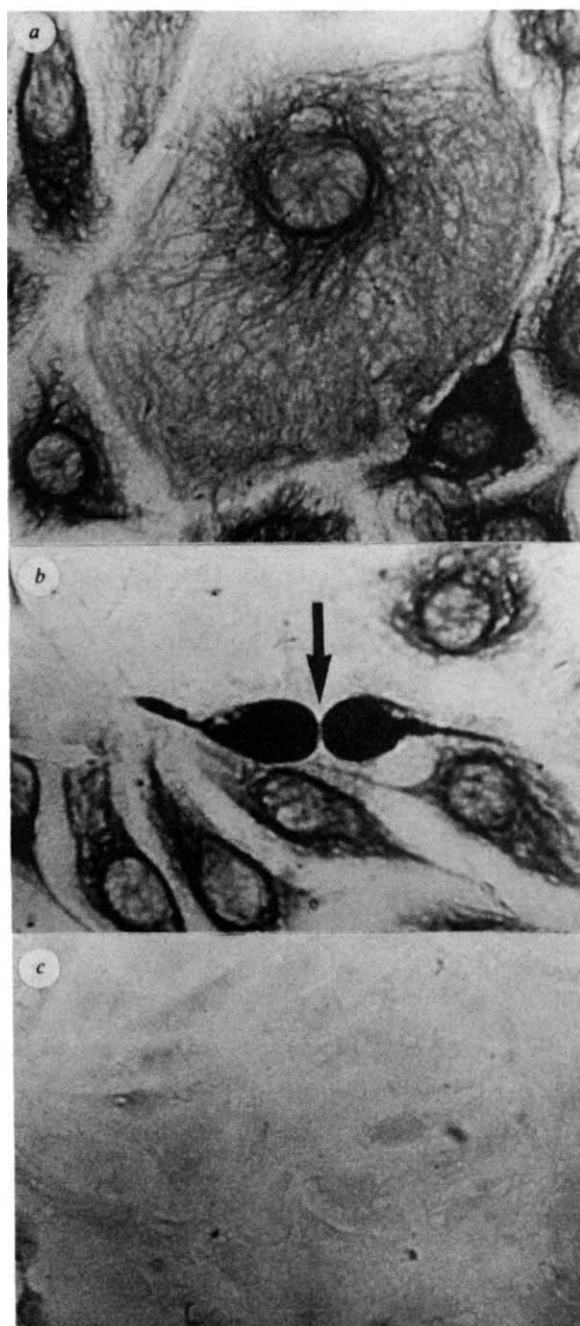


Fig. 1 Micrographs of JTC-12 cells fixed in acetone and stained for GC by the immunoperoxidase method. Peroxidase-labelled anti-rabbit immunoglobulin raised in goats (E-Y Laboratories) diluted 1:30 was used as secondary antibody. As control, antibody preabsorbed with GC was substituted for the primary anti-GC antibody. Briefly, antibody solution was mixed with liposomes consisting of approximate amounts of GC-*lecithin*-cholesterol (1:4:10 by wt) which was prepared by the method previously reported²⁴, and then kept at 37 °C for 1 h with swirling and then at 4 °C overnight. The mixture was centrifuged for 30 min at 12,000 r.p.m. and the supernatant was used as the preabsorbed antibody. *a*, Well developed cytoplasmic fibres stained for GC are demonstrated in a cell with abundant cytoplasm. *b*, Intercellular bridge (arrow) is not stained for GC. *c*, By preabsorption with GC, staining reaction is completely lost from cells. $\times 1,300$.

skeleton of JTC-12 cells. Air-dried, unfixed cells were similarly stained with anti-GC antibody, thus excluding the possibility that cytoskeletal structures visualized with this antibody are artefacts due to the conditions of fixation.

To determine the type of cytoskeleton with which GC is associated, we examined the effect of pretreating the cells with colchicine, cytochalasin B and cold temperature, as described previously^{8,11}. On pretreatment with colchicine or exposure to low temperature, cytoplasmic fibres were completely lost (Fig. 3*a, b*). Colchicine treatment occasionally induced diffuse weak staining all over the cell, but never caused formation of a perinuclear coil and a loop around the cytoplasm as reported for intermediate filaments^{12,13}. In contrast, brief treatment with cytochalasin B, which destroys microfilaments^{14,15}, caused no loss of the fibrous structure stained for GC (Fig. 3*c*). The results of these pretreatments agree with those described for microtubules. A comparison of the staining reactions to anti-GC and rabbit anti-porcine brain tubulin antibodies after pretreatment of the cells with a chloroform-methanol solvent mixture, which is very effective for extraction of lipid, clearly indicated that anti-GC antibody does not react with tubulin itself, as this lipid solvent abolished the staining reaction with anti-GC antibody, but had little effect on that with anti-tubulin antibody (data not shown).

Because it is generally accepted that membrane glycolipids of different cell types have a characteristic composition, in contrast to the ubiquitous distribution of phospholipids^{1,16}, we examined four other epithelial cell lines—HeLa, MDCK, Vero and 8999c, a cell line derived from Morris hepatoma¹⁷—for labelling with anti-GC antibody. Cytoplasmic fibres principally similar to

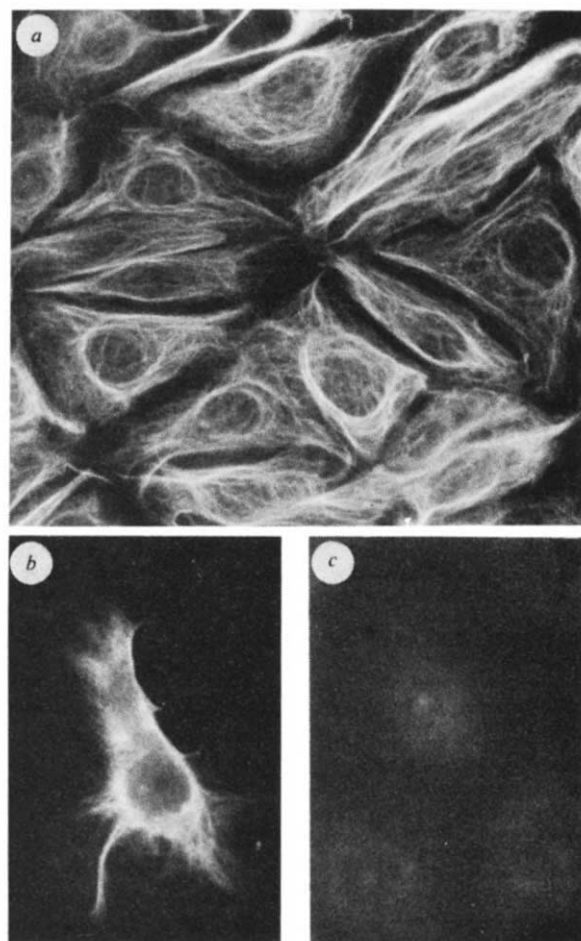


Fig. 2 Immunofluorescence micrographs of JTC-12, HeLa and 8999c cells stained for GC. Fluorescein-conjugated goat antiserum to rabbit immunoglobulin (Hyland) diluted 1:20 was used as secondary antibody. *a*, JTC-12 cells pretreated with non-ionic detergent containing 0.5% Triton-X 100 for 8 min at 37 °C, and then fixed in acetone. *b*, HeLa cells, positive stained. *c*, 8999c cells, negatively stained. $\times 1,300$.

those in JTC-12 cells were seen in HeLa and MDCK cells, but not in Vero or 8999c cells (Fig. 2b, c). To obtain chemical confirmation of this morphological evidence for the presence of GC in the cells examined, glycosphingolipids were isolated from the cells and examined by tritiation and autoradiofluorography. As shown in Fig. 4, the existence of GC in JTC-12 and HeLa cells were clearly demonstrated, but immunocytochemically negative 8999c cells were found to contain only glucocerebroside in the conditions used. MDCK cells have been reported to contain GC as well as glucocerebroside¹⁸.

The present data provide the first evidence for the presence of GC in the microtubule-like cytoskeleton, although definite confirmation of this will require unequivocal isolation and immunoelectron microscopic demonstration of GC bound to microtubules. As suitable conditions for immunocytochemical detection of cytoplasmic glycosphingolipid in cultured cells are now known, it should be possible to examine whether other membrane glycolipids are also connected to microtubules or are part of them. In this context, it is of interest that galactocerebroside¹⁹, galactocerebroside III sulphate (sulphatide)²⁰⁻²² and ganglioside²³ occur in the cytosol of brain as a glycolipid-protein complex and that the pattern of the molecular composition of cytosolic ganglioside is similar to that of membrane-bound ganglioside²³. If the association of glycosphingolipid with microtubules is a general phenomenon, this may pose several important problems with regard to the internalization mechanism of cells, receptor-mediated endocytosis, the mechanism of transmitting a biological message received at the plasma membrane to the nucleus, the transport mechanism of glycolipids from their site of synthesis to the cell surface and other cytoskeleton-related cellular activities.

We thank Dr I. Yahara, Tokyo Metropolitan Institute of Clinical Medicine, for providing rabbit anti-tubulin antibody.

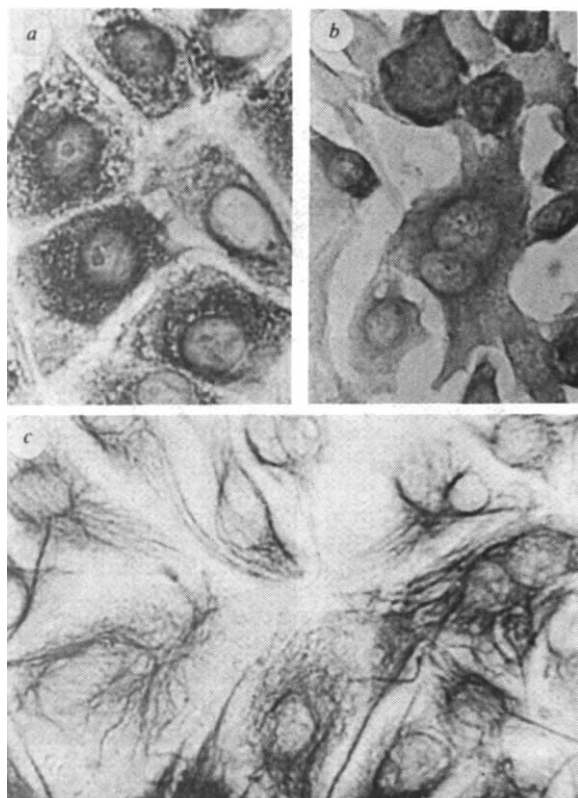


Fig. 3 Micrographs of JTC-12 cells stained for GC by immunoperoxidase method after various pretreatments. *a*, Pretreated with $10 \mu\text{g ml}^{-1}$ colchicine (Sigma) for 18 h at 37°C . *b*, With cold according to the method of Weber *et al.*⁸. *c*, With $5 \mu\text{g ml}^{-1}$ cytochalasin B (Aldrich) for 1 h at 37°C . $\times 1,300$.

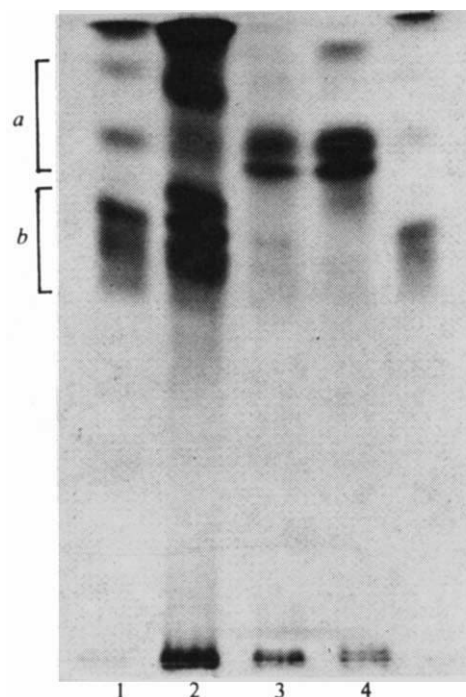


Fig. 4 Autoradiofluorography of tritiated neutral glycosphingolipids of JTC-12, HeLa and 8999c cells. Lipids were extracted from 10^7 cells with 5 ml each of chloroform/methanol/water (30:10:1 by vol.), (10:10:1 by vol.) and (10:20:3 by vol.), respectively. The extracts were combined, evaporated to dryness and submitted to the Folch-Pi partition procedure²⁵. The lower chloroform phase was evaporated, dried completely over P_2O_5 , and peracetylated with pyridine/acetic anhydride (2:1 by vol.) at 50°C for 24 h. The peracetylated glycosphingolipids were isolated by the method of Saito and Hakomori²⁶. After deacetylation, glycosphingolipids were tritiated with potassium boro- ^{3}H -hydride ($616 \text{ mCi mmol}^{-1}$; Amersham) in the presence of palladium chloride as catalyst^{27,28}. ^3H -labelled glycosphingolipids of each cell line (10^6 cells) were applied to a borate-impregnated silica gel thin-layer plate (Merck). ^3H -labelled galactocerebroside and ^3H -glucocerebroside were separated by developing the plate with chloroform/methanol/water/15M ammonia (280:70:60:1 by vol.)²⁹. After development, ^3H -labelled glycosphingolipids were located by autoradiofluorography as described previously²⁸. *a*, Glucocerebroside; *b*, galactocerebroside; 1 and 5, ^3H -glucocerebroside and ^3H -galactocerebroside of bovine brain as reference standard; 2, JTC-12; 3, HeLa; 4, 8999c.

This study was supported in part by grants from the Ministry of Education, Science and Culture of Japan, a Research Grant for Intractable Diseases from the Ministry of Health and Welfare of Japan, and a grant from the Mitsubishi Foundation.

Received 23 April; accepted 13 July 1981.

1. Yamakawa, T. & Nagai, Y. *Trends biochem. Sci.* **3**, 128-131 (1978).
2. Sakakibara, K., Iwamori, M., Uchida, T. & Nagai, Y. *Experientia* **37**, 772-774 (1981).
3. Takaoka, T., Katsuta, H., Endo, M., Sato, K. & Okumura, H. *Jap. J. exp. Med.* **32**, 351-368 (1962).
4. Ishizuka, I., Tadano, K., Nagata, N., Niimura, Y. & Nagai, Y. *Biochim. biophys. Acta* **541**, 467-482 (1978).
5. Sakakibara, K. & Tsukada, Y. *Gann* **71**, 679-685 (1980).
6. Uchida, T. & Nagai, Y. *J. Biochem.* **87**, 1839-1841 (1980).
7. Sternberger, L. A. in *Immunocytochemistry* 2nd edn, 354 (Wiley, New York, 1979).
8. Weber, K., Pollack, R. & Bibring, T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 459-463 (1975).
9. Gordon, W. E., Bushnell, A. & Burridge, K. *Cell* **13**, 249-261 (1978).
10. Osborn, M. & Weber, K. *Cell* **12**, 561-571 (1977).
11. Weber, K., Pathke, P. C., Osborn, M. & Franke, W. W. *Exp. Cell Res.* **102**, 285-297 (1976).
12. Lazarides, E. *Nature* **283**, 249-255 (1980).
13. Hynes, R. O. & Destree, A. T. *Cell* **13**, 151-164 (1978).
14. Henderson, D. & Weber, K. *Exp. Cell Res.* **134**, 301-316 (1979).
15. Karsenti, E. *et al. J. Histochem. Cytochem.* **26**, 934-947 (1978).
16. Hakomori, S. *Biochim. biophys. Acta* **417**, 151-163 (1975).
17. Kido, H., Sanada, Y., Katsunuma, N. & Morris, H. P. *Gann* **68**, 691-696 (1977).
18. Tadano, K. & Ishizuka, I. *Proc. Japan Conf. Lipids* **22**, 353-357 (1980).
19. Yahara, S., Singh, I. & Kishimoto, Y. *Biochim. biophys. Acta* **619**, 177-185 (1980).
20. Jungalwals, F. B. *J. Lipid Res.* **15**, 114-123 (1974).
21. Herschkowitz, N., McKhann, G. M., Saxena, S. & Shooter, E. M. *J. Neurochem.* **15**, 1181-1188 (1968).
22. Herschkowitz, N., McKhann, G. M., Saxena, S., Shooter, E. M. & Herndon, R. J. *Neurochem.* **16**, 1049-1057 (1969).

23. Sonnino, S., Ghidoni, R., Marchesini, S. & Tettamanti, G. *J. Neurochem.* **33**, 117–121 (1979).
24. Nagai, Y. & Ohsawa, T. *Jap. J. exp. Med.* **44**, 451–464 (1974).
25. Folch, J., Lees, M. & Sloane-Stanley, G. H. *J. biol. Chem.* **226**, 497–509 (1957).
26. Saito, T. & Hakomori, S. *J. Lipid Res.* **12**, 257–259 (1971).
27. Schwarzmann, G. *Biochim. biophys. Acta* **259**, 104–114 (1978).
28. Momoi, T. & Wiegandt, H. *Hoppe-Seyler's Z. physiol. Chem.* **361**, 1201–1210 (1980).
29. Kean, E. L. *J. Lipid Res.* **7**, 449–452 (1966).

Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences

Monika Lusky & Michael Botchan

Department of Molecular Biology and Virus Laboratory,
University of California, Berkeley, California 94720, USA

SV40 DNA replication requires interactions between proteins provided by the host cell, the virus-coded A protein and a viral sequence of ~80 base pairs (bp), which serves as an origin of DNA replication^{1,2}. The apparent simplicity of this virus has provided a useful means of producing^{3,4} SV40 virions that carry specific eukaryotic genes which can infect cells in culture and replicate to a high copy number. One disadvantage to this approach is that the viral DNA sequences necessary for replication and the gene of interest must be precisely tailored to package efficiently into SV40 virus particles. In an attempt to circumvent this limitation, we have studied the replication of SV40–pBR322 plasmids after transfection into simian cells with the ultimate goal of using the viral replicon as a means of amplifying the expression of cloned fragments inserted in these vectors. Such a dual-host replicon also offers the possibility of rescuing and subsequently cloning genes which are properly expressed and stably integrated into the cell genome⁵. We and others (refs 5–8 and Schaffner, personal communication), have noted that such plasmids propagated in *Escherichia coli* replicate poorly if at all after transfection of simian cells. Furthermore, recombinant plasmids isolated from the transfected simian cells subsequently show a reduced ability to retransform *E. coli*⁵. We now demonstrate that both these phenomena are related to the presence of a *cis*-acting DNA sequence in the bacterial vector, pBR322. SV40–pBR322 plasmids lacking this sequence replicated as effectively as authentic SV40 DNA in simian cells and were re-established as plasmids in *E. coli* as effectively as plasmid DNAs extracted from *E. coli*.

We have studied the replication of recombinant plasmids in cultured simian cells after their transfection by the DEAE-dextran technique⁹. At various times after transfection, low molecular weight DNA was isolated by the method of Hirt¹⁰, fractionated by electrophoresis in agarose gels and further analysed by Southern blot hybridization¹¹. We used the increase in supercoiled DNA with time as a measure of DNA replication and CV-1 cells and COS7 cells as the recipient simian cells. COS7¹², a permissive line of CV-1 cells transformed by origin-defective SV40 DNA, synthesizes a viral A gene product capable of complementing A gene mutants of SV40¹².

As preliminary experiments indicated that a significant fraction of the input DNA persisted throughout the time course and that only a small fraction of this material became part of the replicating pool (see Fig. 1 legend), it was necessary to transfect small amounts of DNA to observe replication of the input (10 ng DNA per 60-mm dish). Figure 1 clearly shows that a variety of pBR322–SV40 recombinant DNAs, different in size and orientation, did not replicate to any significant level in CV-1 cells and that their replication rate in COS7 cells was markedly lower than that of wild-type SV40 DNA. This lower replication rate does not seem to be a reflection of their size, for although the plasmid pOK3 is the same size as wild-type SV40 DNA, it is replicated at the same reduced rate as pOT, a plasmid 2.5 kilobases (kb) larger than SV40 DNA (Fig. 1). Note that both these recom-

binant plasmids contain an SV40 origin of replication and an intact A gene.

These experiments do not indicate whether SV40 DNA propagated in *E. coli* is capable of normal A gene expression in simian cells. Two types of experiment eliminate this consideration. First, SV40 DNA excised from recombinant plasmids by digestion with enzymes which cleave pBR322 many times (such as *TacI*) yielded intact viral DNA which replicated as wild-type SV40 DNA. Moreover, pDAX (Fig. 1) and pJY1 (Fig. 2) were cleaved with the restriction enzyme *Bam*HI to liberate intact linear SV40 DNA from the pBR322 moiety. Transfection of these mixtures into CV-1 and COS7 cells resulted in replication equivalent to that of SV40 DNA^{6,13} (data not shown). Second, when pDAX, which contains a partial head-to-tail tandem duplication of the viral DNA, was transfected into either CV-1 cells or COS7 cells, excision of genomic length viral DNA evidently occurred *in vivo*, leading to accumulation of detectable form I and II SV40 DNA⁶ (Fig. 1). These results *in toto* suggest that the SV40–pBR322 recombinants replicated poorly due to a *cis*-acting element(s) present in the chimaeric molecules. In this context, the reduced plaque-forming efficiency of unit length SV40 or polyoma–pBR322 recombinants relative to viral DNA may be partly due to their inefficient replication as well as to a low frequency of genomic viral DNA excision (refs 14, 15 and our unpublished observations).

Table 1 summarizes the quantitative differences in accumulation of the various plasmids relative to SV40 DNA, determined using microdensitometer scans of autoradiograms as described in Table 1 legend. The replication rates of the various recombinant plasmids were 20- to 100-fold lower than those of wild-type SV40 DNA transfected into COS7 cells. The differences were larger in CV-1 cells but were not measured.

We also observed that the recombinant plasmids extracted from simian cells transformed *E. coli* 50- to 100-fold less efficiently than the 'same' plasmids extracted from *E. coli* (Table 1). This was surprising because pJY1 (see Fig. 2) and pOT DNAs extracted from *E. coli* cells retransformed bacterial cells to ampicillin resistance at efficiencies equivalent to pBR322. Furthermore, when mixed with Hirt extracts of uninfected simian cells, these DNAs showed no loss of ability to be re-established as plasmids in *E. coli*.

As SV40 replication in simian cells produces defective viral DNA, we reasoned that if small quantities of rearranged plasmids undetected by blot hybridization had in fact been created by replication in COS7 cells, some of these rearrangements might have deleted segments responsible for one or both of the phenomena discussed above. We therefore purified plasmid DNA from over 100 *E. coli* cells transformed with pJY1 DNA

Table 1 Transformation of *E. coli* HB101 by plasmids extracted from COS7 cells

DNA	Transformants per µg DNA	Relative accumulation in COS7 cells
(1) SV40	NA	1.1
(2) pBR322	3.7×10^6	0
(3) pBR322	2.2×10^6	0
(4) pJY1	$3.9-7.8 \times 10^4$	0.01–0.05
(5) pOT	$2.2-3.8 \times 10^4$	0.02–0.05
(6) pJYM	$1.8-2.4 \times 10^6$	1–1.1

Subconfluent monolayers of COS7 cells were transfected with recombinant plasmids and the low molecular weight DNAs extracted 48 h later. To quantitate the amounts of plasmid DNA present in these extracts, samples of the Hirt supernatants were linearized by restriction enzyme digestion and subjected to Southern analysis along with standards of 1, 3.2, 10 and 32 ng of purified plasmid DNAs. The autoradiograms were scanned using a Joyce-Loebel microdensitometer. The sample DNA concentrations were determined from the areas under the densitometric curves. Within each experiment the film response was linear with respect to the standards, and the experimental point fell within the calibrations. We then used equal quantities to transform *E. coli* HB101, selecting for Amp^r colonies. Line 2 shows the transformation efficiency for pBR322 propagated in *E. coli*; line 3 shows a mixing experiment in which uninfected COS7 cell extract was mixed with pBR322; lines 4–6 show the transformation activities for recombinant plasmids pJY1, pOT and pJYM all extracted from COS7 cells. The two numbers given for each line of results show the range of experimental values obtained from several such experiments. NA, not analysed.

extracted from COS7 cells and examined them for molecules with either a selective transforming ability or a higher replication rate in simian cells. Restriction analysis revealed that 10 of these plasmid DNAs had sustained overlapping deletions near the pBR322 origin of replication. The detailed structures of these deletions will be described elsewhere. Of the 10 deletion derivatives isolated, 5 replicated in both COS7 and CV-1 cells as intact plasmids at rates equal to that of authentic SV40 DNA. Figure 2 compares the replication of one such plasmid, pJYM, with that of the parental plasmid pJY1 and SV40 DNA. This result shows directly that certain pBR322 DNA sequences interfered with SV40 replication in the eukaryotic cells. Detailed restriction mapping and DNA sequence analysis revealed that the specific deletion of pBR322 DNA in pJYM spans the region between nucleotide positions¹⁶ 1,092 and 2,485 (see Fig. 2). All plasmids with deletions between the pBR322 origin of replication and *Pvu* II site had the replication properties of pJYM, whereas those plasmids such as pOK3 which retain plasmid DNA beyond the *Pvu* II site intact (see Fig. 1) replicated poorly. We conclude that the 'poison' in the plasmid vectors lies between the *Pvu* II site and the pBR322 origin of DNA replication. To our surprise, the pJYM-like vectors, whether extracted from CV-1, COS7 or *E. coli* cells, retransformed *E. coli* with equal efficiency (see Table 1). These results imply that the two restrictions described were caused by DNA sequences which either overlap or are very close to each other in pBR322.

Interestingly, these studies showed that bona fide SV40 DNA replicates more efficiently in COS7 cells than in CV-1 cells (see Fig. 1). This finding may be explained by proposing that A protein levels are not saturating in lytically infected CV-1 cells and that the additional amount of A protein present in COS7

cells results in a greater viral replication. Consistent with this notion, Myers and Tjian⁸ have described plasmids which contain a 311-bp fragment (*Eco*RII G) spanning the SV40 viral origin of replication, and in confirmation of their results we have found that pBR322 containing three copies of this fragment will replicate after its transfection into COS7 cells (see Fig. 1). However, after transfer of the origin fragment to pML-1, a plasmid vector deleted in the inhibitory sequences, derived from pJYM by removing the *Bam*HI insert of SV40, the replication rate of this construction (called pML-RII G) was not equivalent to pJYM in COS7 cells. Furthermore, there was only a fivefold difference in accumulation between pML-RII G and pSV-01, a similar recombinant maintaining the pBR322 inhibitory sequences⁸. Clearly, these results and those summarized in Table 1 differ in that the replicating molecules here do not contribute to the total A protein pool within COS7 cells.

We do not know the molecular basis for the restrictions described. The control of replication rates of SV40 DNA has generally been thought to occur at the level of initiation; in the present situation it seems unlikely that a sequence >3 kb distant from the viral origin would interfere with this. We believe processes other than initiation may be regulated. Hanahan *et al.*⁵ have proposed that DNA modifications produced in the simian cell (such as DNA methylation) may influence replication elongation rates. These same modifications may also influence a plasmid's ability to be re-established in *E. coli*. In support of this speculation, the primary DNA sequence of the bulk of the pBR322-SV40 recombinants extracted from the simian cells has not been altered, although the molecules have reduced abilities to be re-established in *E. coli*. In this regard, we have not found general methylation of plasmid *Hpa*II sites, although other site-specific modifications cannot be excluded. The poison

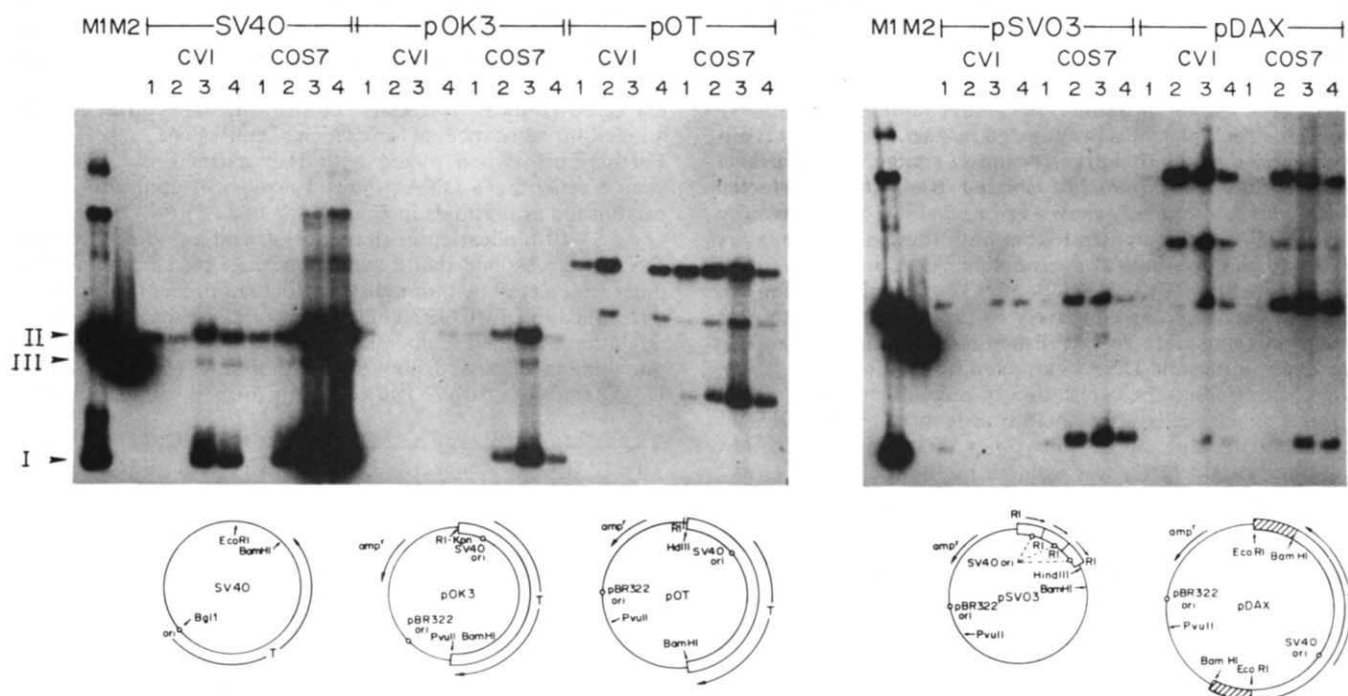


Fig. 1 Analysis of DNA replication after transfection of simian cells. 10 ng of either virion or recombinant plasmid DNAs were applied to subconfluent 60-mm monolayers of CV-1 or COS7 cells. At 2, 24, 48 and 60 h post-transfection, the cells were subjected to the Hirt extraction protocol as previously described¹⁸. One-fifth of the total aliquot from each DNA sample was applied and separated by electrophoresis through 0.8% agarose gels. The gels were treated with HCl (0.2 M) before transfer to nitrocellulose and the DNAs were detected by hybridization with ³²P nick-translated PJY1 probe (specific activity 1–2 × 10⁸ d.p.m. per μg, 10 ng per ml final concentration in hybridization mix). Kodak XR film was applied to the filters for 12 h with a single intensifying screen at –70 °C. The numbers above each slot refer to the different time points. The structures of the plasmids are shown below the autoradiogram. pOT, a 7.4-kb recombinant plasmid described previously⁵, contains an entire SV40 early region and origin of replication as does pOK3, a 5.2-kb plasmid. pOK3 contains continuous SV40 DNA encompassed between the viral *Bam*HI site to its *Kpn*I site; this fragment is linked to pBR322 through the plasmids' *Eco*RI and *Pvu*II sites. pSV03 is a derivative of pSV01⁸ and contains three tandem copies of the SV40 *Eco*RII G fragment. pDAX, a 10.2-kb plasmid, contains the entire SV40 DNA plus a tandem duplication of the viral sequences spanning the *Bam*HI site to the *Eco*RI as indicated. The slot labelled M1 contains 10 ng of SV40, M2 contains 10 ng of SV40 DNA linearized by cleavage with *Bam*HI. It is clear that by 2 h after application to the cells the bulk of the 10 ng applied to the cells has been lost. The DNA becomes further degraded, but by 48 h post-transfection the net amount of virus-specific DNAs extracted from the cultures exceeds the input by 5–10-fold (see also Fig. 2). If microgramme quantities of DNAs are used, 'persisting' DNA obscures these dynamics. Most of the persisting DNA does not enter the replicating pool as measured by other replication assays⁶. If we assume that 1% of the cells actually take up the DNA in a biologically significant manner, the copy numbers for SV40 DNA per cell are ~2 × 10⁶ per cell for COS7 cells and 400,000 copies per cell for CV-1 cells.

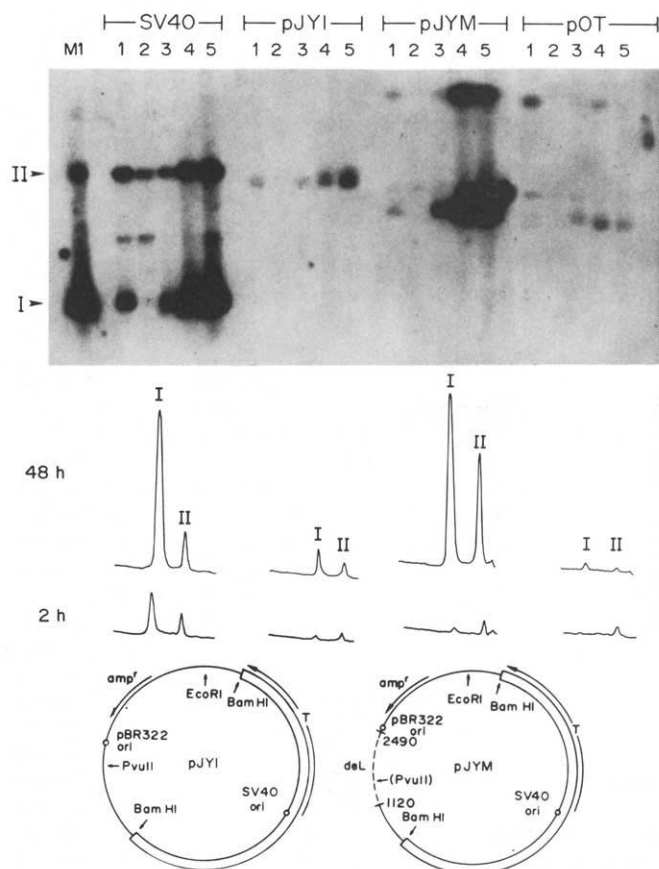


Fig. 2 Comparisons of replication of recombinant plasmids pJY1, pJYM and pOT to SV40 DNA. Experimental details are described in Fig. 1 legend. After transfection, low molecular weight DNAs were extracted at 2, 12, 24, 36 and 48 h; the numbers above each slot refer to these time points. The 10-ng marker of SV40 (forms I and II) DNA was applied to the slot marked M1. Microdensitometer traces of a shorter exposure of the autoradiograms from the 2-h and 48-h time points are shown. pJY1 is a recombinant plasmid consisting of a full-length *Bam*HI SV40 linear sequence inserted into the pBR322 unique *Bam*HI site. pJYM was derived from pJY1 as described in the text; the dotted lines in the drawing of the pJYM structure indicate the boundaries of deleted pBR322 sequences.

sequence has been defined to within a 430-bp segment which overlaps the *nic/bom* site in pBR322¹⁷, a *cis*-acting sequence involved in mobilization of ColE1-type plasmids. Proteins encoded by the *mob* genes prepare ColE1 DNA for conjugal transfer presumably by acting at the *nic/bom* site and thus initiating the transfer process. We think that *nic/bom* site modifications of SV40-pBR322 plasmids may occur in the simian cell which influence the establishment of these DNAs in bacterial cells.

We thank Dr Y. Gluzman for supplying the COS7 cell line and for the plasmid pJY-1. Mr Doug Hanahan constructed the pOK3 plasmid in an attempt to define the inhibitory pBR322 DNA sequences. R. Myers and R. Tjian provided us with pSV03, T. Mitchison, while a URP at Cold Spring Harbor, made plasmid pDAX, and M.L. was supported by a grant from Deutsche Forschungsgemeinschaft. This work was funded by grant MV 91 from the American Cancer Society. We also thank Mr Bernard Danovitch for technical assistance and Mr Warren Gish for help in the preparation of the manuscript.

Received 20 March; accepted 23 June 1981.

1. Tegtmeyer, P. *J. Virol.* **10**, 591-598 (1972).
2. Nathans, D. *Science* **206**, 903-909 (1979).
3. Hamer, D. H. & Leder, P. *Nature* **281**, 35-40 (1979).
4. Mulligan, R. C., Howard, B. H. & Berg, P. *Nature* **277**, 108-114 (1979).
5. Hanahan, D., Lane, D., Lipsich, L., Wigler, M. & Botchan, M. *Cell* **21**, 127-139 (1980).
6. Peden, W. C., Pipas, J. M., Pearson-White, S. & Nathans, D. *Science* **209**, 1392-1396 (1980).
7. Benoist, C. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3865-3869 (1980).
8. Myers, R. M. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6491-6495 (1980).

9. McCutchan, J. H. & Pagano, J. S. *J. natn. Cancer Inst.* **41**, 351-356 (1968).
10. Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).
11. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
12. Gluzman, Y. *Cell* **23**, 175-182 (1981).
13. Gluzman, Y., Frisque, R. J. & Sambrook, J. *Cold Spring Harb. Symp. quant. Biol.* **43**, 293-300 (1979).
14. Fried, M. *et al.*, *Nature* **279**, 811-816 (1979).
15. Israel, M. A., Chan, H. W., Rowe, W. P. & Martin, M. A. *Science* **203**, 883-887 (1979).
16. Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2721 (1978).
17. Clark, A. J. & Warren, G. J. *A. Rev. Genet.* **13**, 99-125 (1979).
18. Botchan, M. R., Topp, W. & Sambrook, J. *Cold Spring Harb. Symp. quant. Biol.* **43**, 709-719 (1980).

Site of linkage between adenovirus type 12 and cell DNAs in hamster tumour line CLAC3

R. Deuring, U. Winterhoff, F. Tamanai*, S. Stabel & W. Doerfler

Institute of Genetics, University of Cologne, Cologne, FRG

The patterns of integration of persisting viral DNA have been examined in 30 to 40 different adenovirus type 2 (Ad2)- and type 12 (Ad12)-transformed hamster cell lines, Ad12-induced hamster and rat tumours and tumour cell lines¹⁻⁵. These patterns have proved to be quite complex, and so far no evidence has been found for specific sites of virus integration. However, the methods of analysis used so far, are not sufficiently sensitive to decide definitively the issue of specificity. Moreover, specific signals directing the interactions between viral and cellular DNA may reside in DNA sequences distant from the site of junction between the two genomes and may be hidden in hitherto undetected structural peculiarities of DNA sequences. From the DNA of an Ad12-induced hamster tumour line (CLAC3) carrying five copies of the Ad12 genome¹, the site of junction between the left terminus of Ad12 DNA and cellular DNA has now been recloned in a λ gtWES- λ B' vector. Determination of the nucleotide sequence at this site of junction reveals that the Ad12 DNA sequence is preserved starting with base pair (bp) 46 of the authentic left end of Ad12 DNA. The first 82 bp of the recloned fragment are not viral and contain scattered, patch-like octa- to undecanucleotide pair homologies to distant sequences within the first 2,722 left terminal base pairs of Ad12 DNA. Close to and encompassing the site of junction a stemmed loop can be constructed based on extensive regions of dyad symmetry. The stability of this loop and its biological function are still uncertain.

The total DNA of cell line CLAC3 was cleaved with the restriction endonuclease *Eco*RI; fragments larger than 4.0 kilobases (kb) (*Eco*RI fragment D; Fig. 1a) were selected by velocity sedimentation on sucrose density gradients⁶, and the selected fragments ligated to the 'arms' of bacteriophage λ gtWES- λ B' (ref. 7) which was used as vector. Previously described methods of *in vitro* packaging of recombinant DNA were used^{8,9}. Phage containing DNA sequences derived from the termini of Ad12 DNA were selected by using the terminal *Eco*RI fragments A and C as probes (Fig. 1a). Recombinant phage were purified by repeated cycles of single plaque isolations and propagated to yield large amounts of recombinant DNA which was characterized further (Fig. 1).

Three plaques of recombinant λ phage (clones 35, 38 and 41) containing DNA sequences homologous to the left terminus of Ad12 DNA were isolated, and the DNAs purified from these phage preparations were analysed by restriction endonuclease cleavage, Southern blotting¹⁰ and DNA-DNA hybridization¹¹. Figure 1 summarizes the results of this analysis for the DNA of clone 35. The recombinant λ DNA isolated contains two inserts of 5.4 and 3.0 kb (Fig. 1b-e). The results demonstrate that the

* Present address: The Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

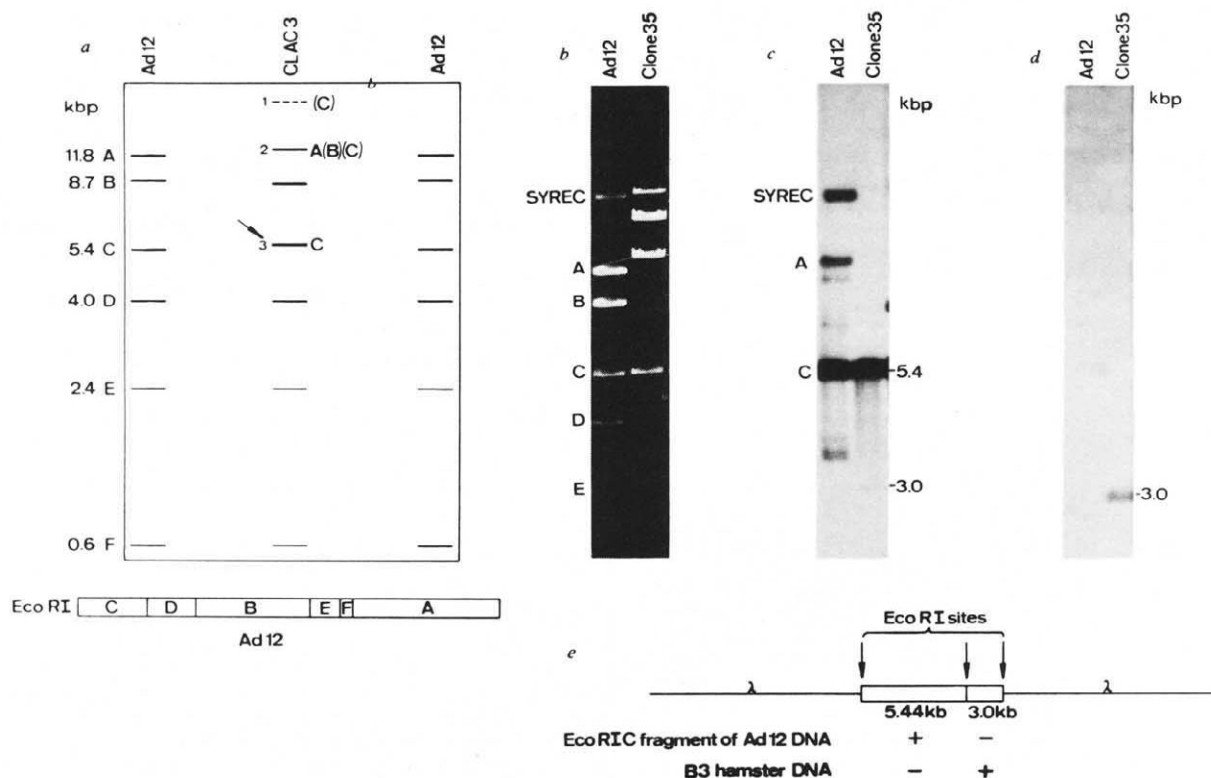


Fig. 1 Analysis of the DNA of clone 35, a recombinant bacteriophage λ DNA preparation containing the left end of Ad12 DNA integrated into the DNA of hamster tumour cell line CLAC3. *a*, Schematic representation of the Ad12-specific DNA fragments in CLAC3 DNA after cleavage with *Eco*RI, Southern blotting¹⁰ and DNA-DNA hybridization¹¹. Individual bands were identified by using purified or cloned Ad12 DNA fragments¹⁹ as hybridization probes, as described previously¹. Authentic Ad12 DNA cleaved with the same enzyme was co-electrophoresed in the same agarose slab gel. The C band designated 3 is off size, that is, of slightly higher molecular weight than the authentic C band of Ad12 DNA. Thus, sequences other than true Ad12 DNA sequences must be attached to it. Band 3 has been cloned in three isolates of recombinant λ phage (clones 35, 38 and 41). The segment cloned and partly sequenced as described in the present report is designated by an arrow. *b*, The DNA of bacteriophage λ gtWES-AB⁷ was cleaved with *Eco*RI, and the terminal fragments (arms) of this DNA were purified by velocity sedimentation on sucrose density gradients⁶. The positions of the λ arms in the gradient fractions were identified by electrophoresing small portions of each fraction in an agarose slab gel. Similarly, the DNA from hamster tumour line CLAC3 was cleaved with *Eco*RI and the fragments were separated by velocity sedimentation on sucrose density gradients. Fragments larger than the *Eco*RI fragment D of Ad12 DNA (see Fig. 1*a*) were identified by electrophoresing portions of each fraction on agarose slab gels using *Eco*RI-cut Ad12 DNA as a size marker and were then pooled. Vector DNA and the selected CLAC3 DNA fragments were ethanol precipitated and resuspended in 50 mM Tris-HCl, pH 7.8, and 0.1 mM EDTA. In a typical ligation experiment, 10–25 μ g of DNA were ligated in a total volume of 120 μ l of 50 mM Tris-HCl pH 7.8, 10 mM MgCl₂, 10 mM dithiothreitol, 0.2 mM ATP, 0.1 mM EDTA containing 0.5–1 unit of ligase from T4-infected *Escherichia coli* (Biolabs) and 6 μ g of bovine serum albumin. In general, twice as much vector DNA was used as DNA to be inserted. The DNA mixture was heated to 50 °C for 10 min before the addition of ligase, ATP and dithiothreitol. The ligation reaction was allowed to proceed for 24 h at 4 °C. Ligase (0.5–1 unit) and 2 μ l of 0.16 mM ATP were again added, incubation was continued for 24 h at 4 °C, and the reaction was then monitored by electrophoresing a small portion of the reaction mixture in an agarose slab gel. The ligated DNA was subsequently packaged *in vitro* into λ phage heads by the method of Hohn and Murray^{8,9} using strain BHB2600 (recA⁺) as plating bacteria. DNA from individual plaques was transferred to nitrocellulose filters²⁰, and plaques containing viral DNA sequences homologous to the left or right terminus of Ad12 DNA (see Fig. 1*a*) were identified by DNA-DNA hybridization¹¹ using the ³²P-labelled²¹ *Eco*RI fragments A and C as a mixed hybridization probe¹. Positive plaques were replated three or four times, and sufficient DNA was prepared from the purified phage to carry out restriction analysis of the recombinant DNA. DNA from isolate 35 or Ad12 marker DNA was cleaved with *Eco*RI, and the fragments were separated by electrophoresis on 0.4% agarose gels. A UV-photograph of the ethidium bromide-stained gel is shown. *c*, The DNA from a gel similar to that shown in *b* was transferred to a nitrocellulose filter¹⁰ and hybridized to the purified ³²P-labelled *Eco*RI fragment C. *d*, Experimental conditions were as described in *c*, except that ³²P-labelled hamster cell DNA (from line B3, a subline of BHK21) was used as hybridization probe. *e*, The scheme depicts the arrangement and derivation of DNA fragments from line CLAC3 that were cloned in λ gtWES-AB DNA. Hybridization of the 5.4- and 3.0-kb fragments to the *Eco*RI C fragment of Ad12 DNA and to B3 hamster cell DNA, respectively, is also indicated.

5.4-kb fragment is predominantly viral (Fig. 1*c,d*) and corresponds in size exactly to that of off size band 3 (that is, of higher molecular weight than authentic band C) in the DNA from line CLAC3 (Fig. 1*a,b*). The 3.0-kb fragment hybridizes to hamster cell DNA and not, or only weakly, to the *Eco*RI C fragment of Ad12 DNA (Fig. 1*c,d*). It is striking that all three recombinant clones isolated contain fragments of identical size. Perhaps, in the cell, the 3.0-kb fragment is adjacent to the 5.4-kb fragment. In many cloning experiments using the λ gtWES-AB⁷ vector, we have observed that inserts of ~8 kb apparently constitute a preferred size for this particular vector.

The 5.4-kb fragment from clone 35 was further analysed by excision from the vector with *Eco*RI and subsequent cleavage of DNA with *Msp*I. The fragments generated were compared with those in the authentic *Eco*RI fragment C of Ad12 DNA (Fig. 2) which was also cleaved with *Msp*I. It is apparent that all Ad12-specific fragments of 2.46 kb, ~1.5 kb, 0.55 kb, 0.50 kb,

0.225 kb, 0.143 kb and 0.08 kb (ref. 12 and Fig. 2) in the authentic *Eco*RI fragment C are also present in the 5.4-kb fragment; however, the terminal 0.143-kb fragment is displaced to the position of about 0.18 kb (Fig. 2). We conclude that the left terminal *Eco*RI C fragment in the inserted Ad12 DNA is still largely colinear with the virion DNA fragment, and that DNA sequences have been attached at the terminus (scheme in Fig. 2).

The 5.4-kb fragment excised from clone 35 was terminally labelled and cleaved with the *Msp*I restriction endonuclease, and the nucleotide sequence of the fragment comprising the site of junction (see Fig. 2) was then determined (see Fig. 3 legend for details of the experimental approach). The sequence for both strands was determined by the Maxam-Gilbert method^{13,14} up to position 87 of authentic Ad12 DNA which corresponds to position 124 of the cloned DNA. The sequence to position 143 of Ad12 DNA, that is, position 180 in clone 35, was determined for only one strand as indicated in the scheme of Fig. 3*a*. The

Table 1 Apparent patch homologies between sequences in the first 82 nucleotide pairs of clone 35 (5.4-kb fragment) and the left terminal 2,730 nucleotide pairs of Ad12 DNA

First nucleotide in string (clone 35)	Strand (left terminus)	Direction	String	First nucleotide in string (Ad12) (on 5' → 3' strand)	Frequency of occurrence*
Octanucleotides					
20	5' → 3'	→	CTTGCACT	1,587	4 times
28	5' → 3'	→	GGGGAAAC	353	8 times
43	5' → 3'	→	CTAGACCT	1,732	2 times
60	5' → 3'	→	TGCTTTTT	1,818	21 times
70	3' ← 5'	←	TTGAAAAA	2,722	13 times
40	3' ← 5'	←	GTCCTGTT	1,144	20 times
22	3' ← 5'	←	AAGCATTT	1,292	3 times
Nonanucleotides					
66	5' → 3'	→	TTCAAGAGG	1,309	5 times
38	3' ← 5'	←	CCTGTTTCC	1,146	5 times
Undecanucleotide					
18	3' ← 5'	←	ATTTTAGCTGC	709	

The nucleotide homologies were determined by searching the sequences with a Zilog computer. The polarities of the strands relative to the left terminus of the sequence presented in Fig. 3 and the direction of reading of each string of nucleotides on the corresponding strand are presented. The corresponding homologies in the Ad12 DNA sequence are all located on the 5' → 3' strand and read from left to right.

*The numbers in this column indicate the frequency of occurrence of each string among 440,000 randomly selected nucleotide pairs in pro- and eukaryotic DNA sequences.

sequence is flanked by an *EcoRI* site on the left (nucleotide pairs 1–5, one G is missing = position 0) and by an *MspI* site on the right (nucleotide pair 180). Nucleotide pair 83 corresponds to nucleotide 46 in authentic Ad12 DNA¹⁵. From there on, the sequence is identical to that of Ad12 DNA at least up to nucleotide pair 143 and also beyond (Fig. 2). Thus, the integrated Ad12 DNA at the left terminus is colinear with virion

DNA starting at nucleotide pair 46. The first 82 nucleotide pairs from the left end of the 5.4-kb fragment of clone 35 are, however, quite different from the authentic left terminus of Ad12 DNA. Peculiar, patch-like regions occur in the sequence within the first 82 nucleotide pairs of the clone that are homologous to sequences within the left terminal 2,722 nucleotide pairs of authentic Ad12 DNA. Table 1 lists homologous octa-, nona- and undecanucleotide sequences; some of these are inverted (3' ← 5' strand in the clone), others are colinear with Ad12 DNA. Table 1 also indicates the frequencies of occurrence of the octa- and nonanucleotide sequences in randomly selected sequences of prokaryotic and eukaryotic origins comprising a total of 440,000 bp. We did not find the undecanucleotide sequence, which contains an *AluI* site (5'-AGCT-3'), among these sequences. The biological significance of these patch homologies and the extent to which they may direct the insertion event are unknown.

Previous work has shown extensive nucleotide sequence homologies at the left termini between nucleotides 1 and ~51 of the DNAs of adenovirus types 5, 7 and 12 (refs 16, 17). This relatively A + T-rich region is also thought to contain one origin of adenovirus DNA replication¹⁶. It is exactly that part of the Ad12 genome that has been replaced in the integrated state by a cellular DNA sequence exhibiting an array of homologies to distant segments of the Ad12 genome. Thus, the authentic viral origin of DNA replication has been replaced. Perhaps this is why the integrated Ad12 genomes cannot replicate under their own control. Computer analysis revealed regions of dyad symmetry with only a few mismatched bases between nucleotide pairs 35 and 54 and nucleotide pairs 71 and 90 (Fig. 3a). These regions of symmetry straddling the site of the junction permit the construction of a stemmed loop, the stem comprising 15 matched and 4 unmatched nucleotide pairs, and the loop containing 16 nucleotides (Fig. 3b). The free energy of this structure amounts to $\Delta G = -13.1$ kcal as determined by the method of Tinoco *et al.*¹⁸. Moreover, sequences of nucleotides between positions 75 and 101 and from 108 to 135 exhibit 70% homology (Fig. 3b). The possible structural arrangements at many more sites of junction between Ad12 DNA and cellular DNA must be determined before any conclusions can be drawn about the functional significance of the loop structure or the observed patch homologies.

Recently, sites of junction have been recloned successfully in our laboratory from three additional tumour or transformed cell lines—lines CLAC1 (ref. 1), H313 (ref. 5) and the Ad2-transformed cell line HE5 (ref. 4) (S. Stabel, R. Neumann, R. Gahlmann, L. Vardimon and W. Doerfler, unpublished results). We consider it unlikely that patch-type homologies to Ad12 sequences indicate a 'scrambling' of viral sequences which could have occurred during cloning because (1) the same sizes of insert were found in three isolates of λ phage clones; (2) the Ad12-

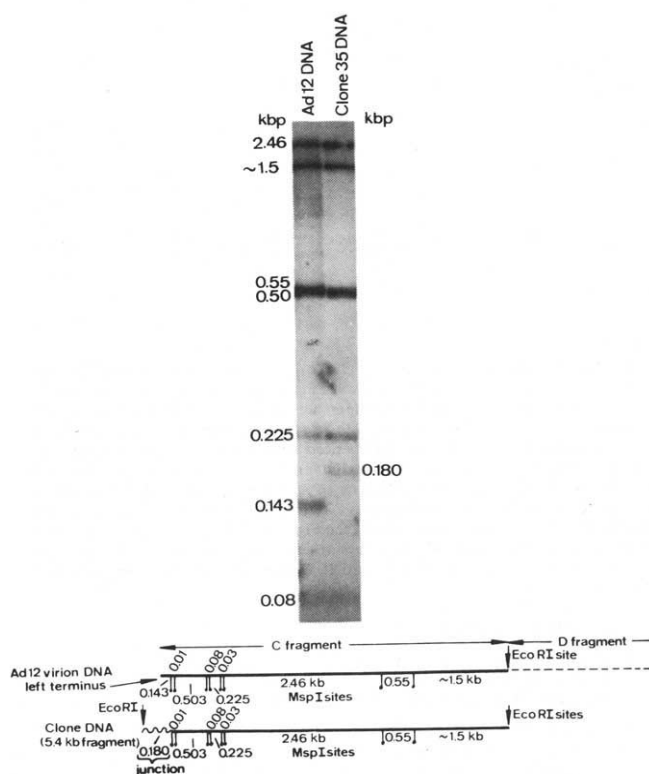
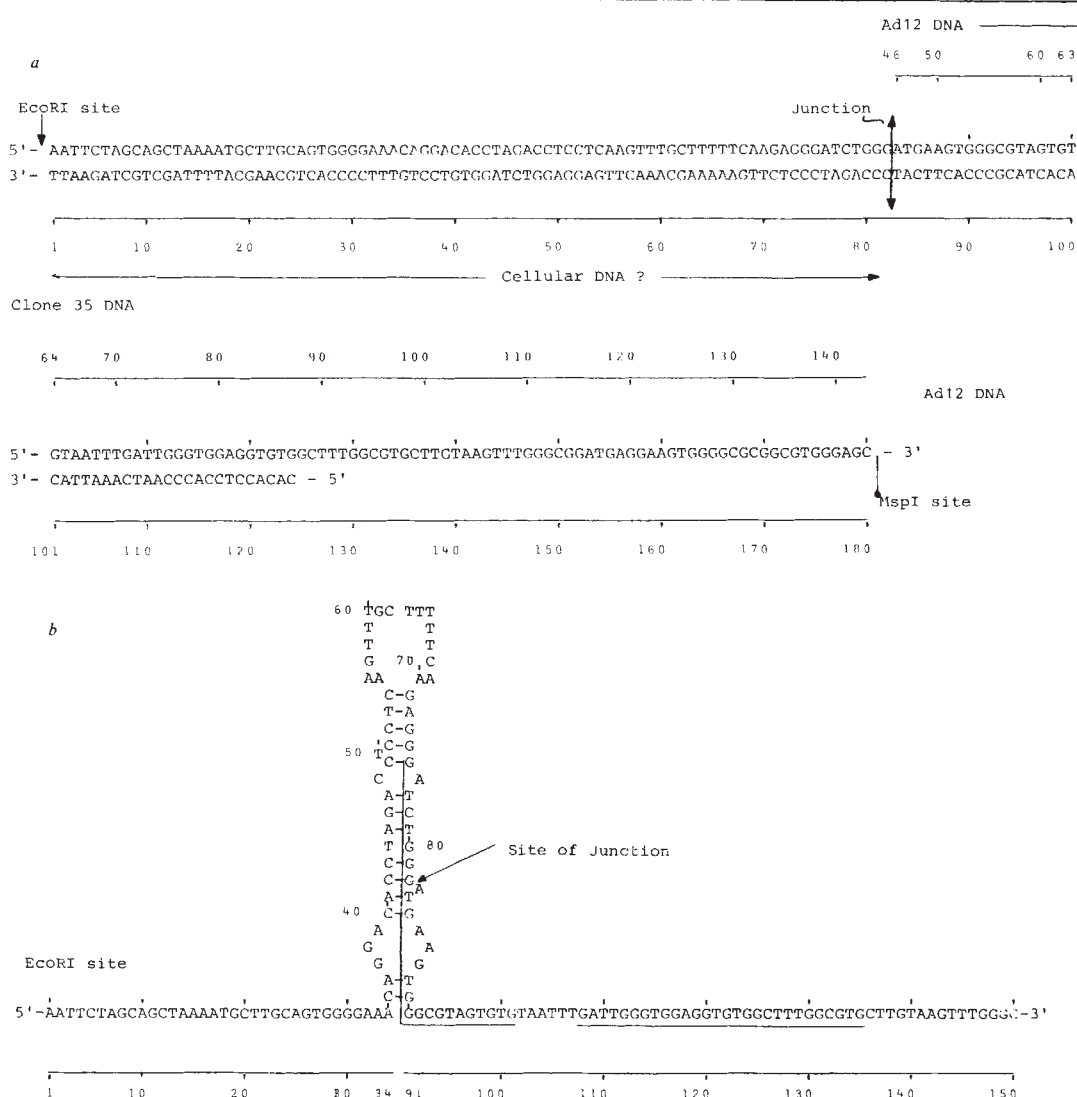


Fig. 2 Restriction enzyme analysis of the 5.4-kb fragment from clone 35. The DNA from clone 35 or from Ad12 virions was cleaved successively with *EcoRI* and *MspI*, and the fragments were separated by electrophoresis on a mixed 0.7% agarose, 5.0% polyacrylamide slab gel²² and transferred to nitrocellulose filters¹⁰. The *EcoRI* fragment C of Ad12 DNA was cloned in the plasmid pBR322 (a gift of A. van der Eb) and was ³²P-labelled by nick translation²¹ before being used as hybridization probe. The sizes in kilobases of the authentic *MspI* fragments at the left end of Ad12 DNA are derived partly from sequence data¹⁵ and partly from the *MspI* map published previously¹². The schematic drawing depicts the *EcoRI* (↓) and *MspI* (↓) sites in authentic Ad12 DNA and the 5.4-kb fragment from clone 35. In part the wavy line indicates adjacent, perhaps nonviral sequences. The terminal fragment of clone 35 (bracketed and designated junction) has been sequenced (see Fig. 3). The length of the off size *MspI* fragment in clone 35 has been estimated to be 175–180 bp.

Fig. 3 Nucleotide sequence analysis. The 5.4-kb fragment from clone 35 was excised from the vector with *EcoRI*. The 5.4-kb fragment was purified by electrophoresis on a 0.5% agarose gel and was terminally labelled using [γ - 32 P] ATP and polynucleotide kinase²³ as described previously¹². Subsequently, the DNA was cleaved with *MspI* and the fragments were separated by electrophoresis on a 6% polyacrylamide gel. The *MspI* map of the left terminus of Ad12 DNA¹² and the analysis presented in Fig. 2 indicated which of the two 32 P-labelled *MspI* fragments corresponded to the left terminus of Ad12 DNA linked to the adjacent DNA sequence at the site of junction. The 5' terminus of this fragment was sequenced by the method of Maxam and Gilbert^{13,14}. The only minor technical modification introduced concerned the time periods allowed for the chemical reactions with pyridine formate (A + G, 10 min, 42 °C), with dimethyl sulphate (G, 1–2 min, 25 °C), with hydrazine alone (C + T, 20–40 min, 37 °C) and with hydrazine in the presence of 1.5 M NaCl (C, 20–40 min, 37 °C). The nucleotide sequence of part of the opposite strand of the terminal fragment was carried out as follows. After excision of the 5.4-kb fragment from clone 35 DNA, the DNA was labelled at its 3' terminus by incubating 1 μ g of DNA in 34 μ l of 135 mM potassium phosphate pH 6.5, 67 mM Tris-HCl pH 7.4, 7 mM MgCl₂, 10 mM 2-mercaptoethanol containing 4 units of the Klenow fragment of DNA polymerase I and 100 μ Ci each of [α - 32 P]-labelled dATP, dGTP and dTTP (Amersham) for 60 min at 14 °C. Subsequently, 5 μ l of an 0.1 mM solution of each of the four unlabelled deoxyribonucleoside triphosphates was added and incubation was continued for 60 min at 14 °C. After ethanol precipitation, the DNA was cleaved with *MspI* and the fragment containing the junction site was isolated and subjected to nucleotide sequence analysis. *a*, Nucleotide sequence at the site of junction of the left end of Ad12 DNA in hamster tumour line CLAC3. Position 83 of the sequence presented corresponds to nucleotide 46 (5' $\rightarrow$ 3' strand) from the left end of Ad12 DNA as sequenced by Sugisaki *et al.*¹⁵. The nucleotide sequence was determined in both strands of the sequence presented up to nucleotide 124. Nucleotides 125–180 in the clone were determined in the 5' $\rightarrow$ 3' direction only. The left end of the sequence carries an *EcoRI* site ($\downarrow$), the right end an *MspI* site (at position 180 in the clone), as expected. The scale on the bottom of the sequence refers to nucleotide numbers in the clone. Starting at the site of junction ($\downarrow$), the scale on top of the sequence marks the authentic Ad12 DNA beginning with nucleotide pair 46 (position 83 in clone 35). The string of the first 82 nucleotide pairs in the sequence has been designated 'cellular DNA?', because the 5.4-kb fragment of clone 35 does not seem to hybridize to hamster cell DNA (see Fig. 1). *b*, Stemmed loop structure at site of junction. Computer analysis revealed the structure presented encompassing the site of junction. Nucleotide numbers are indicated for comparison with the lower scale given in *a*. Only half of the cruciform structure is drawn. The underlined sequences of nucleotides (75–101 and 108–135) are 70% homologous.



specific insert still carries the *EcoRI* sites and exhibits exactly the same length as in CLAC3 tumour cell DNA; (3) we have frequently recloned the internal *EcoRI* fragment B of Ad12 DNA (Fig. 1a) from the DNA of several lines of transformed and tumour cells in the λ gtWES- λ B' vector. The cleavage patterns of the *EcoRI* B fragment from transformed and tumour cells and from Ad12 DNA—recloned in both cases in the same vector—with *EcoRI*, *BamHI*, *MspI* and *HpaII* were identical.

Kei Fujinaga provided unpublished sequence data of Ad12 DNA. F.T. was a short-term fellow of EMBO. We thank Barbara Hohn for strains and advice with the λ *in vitro* packaging procedure, Phil Leder for providing the λ gtWES- λ B' vector, Heinz Schaller, and Hans-Joachim Fritz for help with the nucleotide sequencing procedure, Doug Brutlag and Reiner Leisten for help with the computer analysis, and H. Zimmermann for constructing the electrophoresis apparatus. This research was supported by the Deutsche Forschungsgemeinschaft through SFB 74 and by grant IIB5-FA 8381 from the Ministry of Science and Research of the State of Northrhine-Westphalia.

Received 16 March; accepted 1 July 1981.

1. Stabel, S., Doerfler, W. & Friis, R. R. *J. Virol.* **36**, 22–40 (1980).
2. Eick, D., Stabel, S. & Doerfler, W. *J. Virol.* **36**, 41–49 (1980).
3. Ibelgauts, H., Doerfler, W., Scheidtmann, K. H. & Wechsler, W. *J. Virol.* **33**, 423–437 (1980).
4. Vardimon, L. & Doerfler, W. *J. molec. Biol.* **147**, 227–246 (1981).
5. Kuhlmann, I. & Doerfler, W. *Virology* (submitted).
6. Burlingham, B. T. & Doerfler, W. *J. Virol.* **7**, 707–719 (1971).
7. Tiemeier, D., Enquist, L. & Leder, P. *Nature* **263**, 526–527 (1976).
8. Hohn, B. & Murray, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3259–3263 (1977).
9. Hohn, B. *Meth. Enzym.* **68**, 299–309 (1979).
10. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
11. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683–3687 (1979).
12. Doerfler, W. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **44**, 551–564 (1979).
13. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
14. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
15. Sugisaki, H. *et al.* *Cell* **20**, 777–786 (1980).
16. Tolun, A., Aleström, P. & Pettersson, U. *Cell* **17**, 705–713 (1979).
17. van Ormondt, H., Maat, J. & Dijkema, R. *Gene* **12**, 63–76 (1980).
18. Tinoco, I. Jr, Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362–367 (1971).
19. Vogel, S. *et al.* *Gene* **15** (1981).
20. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
21. Rigby, P. W. J., Diekmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
22. Reiser, J., Renart, J. & Stark, G. R. *Biochem. biophys. Res. Commun.* **85**, 1104–1112 (1978).
23. Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **54**, 158–165 (1965).

Thermoproteales—a third order of thermoacidophilic archaeobacteria

Wolfram Zillig, Jenn Tu & Ingelore Holz

Max-Planck-Institut für Biochemie, 8033 Martinsried b. München, FRG

The prokaryotic domain has been divided into the urkingdoms of the eubacteria and the archaeobacteria on the basis of comparison of their 16S rRNA sequences¹⁻³. Differences in their cell walls⁴, membrane lipids⁵ and mechanisms of transcription⁶ and translation⁷⁻¹⁰ provide further evidence for this division. The demarcation of two branches of archaeobacteria, the methanogens, including the extreme halophiles¹¹⁻¹³, in one branch, and the thermoacidophiles, represented by the two orders *Sulfolobus*¹⁴ and *Thermoplasma*¹⁵, on the other, is based on physiology and on a comparison of component patterns of DNA-dependent RNA polymerases¹⁶⁻²⁰ rather than rRNA sequences. We show here that *Thermoproteus tenax*, a novel anaerobic, sulphur-respiring, multiform thermoacidophilic archaeobacterium isolated from solfataric springs of Iceland²¹, represents a third order of the thermoacidophilic branch of archaeobacteria and suggests a specific relationship between thermoacidophilic archaeobacteria and primitive eukaryotes such as yeast.

The RNA polymerase of *Thermoproteus tenax*, like that of *Methanobacterium*¹⁷, is highly oxygen sensitive and was therefore purified under strict exclusion of air using an anaerobic hood²² and the same purification procedure as used for other archaeobacterial RNA polymerases^{6,16-20} (see Fig. 1 legend).

DNA cellulose chromatography²³ yielded three peaks of which the last and largest, eluted in a salt gradient at 0.5 M NH_4Cl in buffer A, represents the RNA polymerase (Fig. 1a). Residual impurities were removed by sucrose glycerol gradient centrifugation (Fig. 1b). With the DNA of *Halobacterium* phage ΦH (ref. 16) the optimal Mg^{2+} concentration was found to be between 10 and 30 mM at 100 mM KCl. The optimal KCl concentration is 100 mM at 30 mM Mg^{2+} . The optimal temperature for ΦH DNA transcription is 86 °C, that for the transcription of poly(dAT.dAT) is 81 °C. However, the enzyme itself is completely stable at even 95 °C, and at 100 °C still has a half life of 135 min. At 85 °C the specific activity with ΦH DNA and T_7 DNA is about 100 mU per mg, that is, of the order of that of *Escherichia coli* RNA polymerase at 37 °C.

As shown in Table 1, the enzyme is insensitive to 100 $\mu\text{g ml}^{-1}$ rifampicin or streptolydigin but is completely inhibited by

Table 1 Transcription properties of DNA-dependent RNA polymerase from *Thermoproteus*

Template	Inhibitor	¹⁴ C-AMP incorporated (pmol per μg enzyme per min)
—	—	2.4
ΦH DNA	—	82.5
ΦH DNA	Rifampicin	76.6
ΦH DNA	Streptolydigin	68.2
ΦH DNA	Heparin	2.3
ΦH DNA	No incubation	0.8
Poly[d(AT) . d(AT)]	—	150.0
Single-stranded calf thymus DNA	—	21.0
<i>Thermoproteus</i> DNA	—	50.9
<i>E. coli</i> DNA	—	71.8
T_4 DNA	—	23.3
T_5 DNA	—	58.1
T_7 DNA	—	95.2
λ DNA	—	66.1
P_2 DNA	—	61.6

Test conditions: total volume 100 μl , substrate concentration 1 mM each, 30 mM MgCl_2 , 100 mM KCl, template 10 μg each, inhibitor 100 μg each, enzyme 3 μg each. Incubation for 10 min at 82 °C.

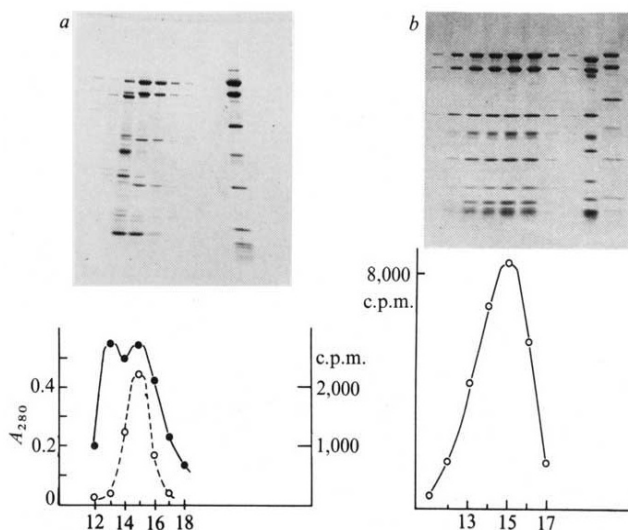


Fig. 1 Purification of DNA-dependent RNA polymerase from *Thermoproteus tenax*. A crude extract was prepared by opening the cells with a French press. The RNA polymerase, together with nucleic acids and many proteins, was precipitated by polyethylenimine (Polyimin P)³¹ and eluted from the precipitate at elevated ionic strength [0.8 M NH_4Cl in buffer A (0.05 M Tris-HCl, 0.01 M EDTA, 0.01 M β -mercaptoethanol, 40% glycerol)]. The enzyme activity was then concentrated by ammonium sulphate precipitation (with 2.5 volumes of saturated $(\text{NH}_4)_2\text{SO}_4$ solution in water). The precipitate was redissolved and dialysed against buffer A containing only 0.022 M NH_4Cl . In heparin cellulose chromatography³², the activity was associated with a shoulder following the main absorbance peak, eluted in a salt gradient at 0.75 M NH_4Cl in buffer A. a, Fractions from DNA cellulose chromatography. Upper part, SDS-polyacrylamide electrophoresis gradient gel (10–25% acrylamide, otherwise according to ref. 33. Right lane: RNA polymerase from *Thermoplasma*. Each lane represents 200 μl of a 9-ml fraction. Lower part: A_{280} (closed circles) and ^{14}C -AMP incorporated in enzyme test on ΦH DNA at 82 °C in optimal conditions for 25 min as described in the text, with 20 μl each of the fractions (open circles). b, Fractions from sucrose glycerol gradient centrifugation. About 1.5 mg of concentrated enzyme from DNA cellulose chromatography, adjusted to 0.5 M NH_4Cl in buffer A containing 10% glycerol, was layered on top of a gradient ranging from 30 to 10% sucrose and from 10 to 5% glycerol in the same buffer and subjected to centrifugation for 34 h at 41,000 r.p.m. in rotor SW 41 of the Beckman Spinco ultracentrifuge. Upper part, electropherogram as in a, with 10 μl each of the fractions. Second lane from right, *Sulfolobus* RNA polymerase; first lane from right, *Thermoplasma* RNA polymerase. Lower part: activity, tested as in a, but with 10 μl each of the fractions.

100 $\mu\text{g ml}^{-1}$ heparin. The time course of transcription has two linear phases, one between 0 and 10 min, and a slower one between 10 and 60 min, the activity depending strongly on the nature of the template (see Table 1).

Both the component pattern and the insensitivity to rifampicin and streptolydigin are characteristic of archaeobacterial

Table 2 Molecular weights ($\times 10^{-3}$) of components of DNA-dependent RNA polymerases from thermoacidophilic archaeobacteria

Designation	<i>Sulfolobus acidocaldarius</i>	<i>Thermoplasma acidophilum</i>	<i>Thermoproteus tenax</i>
A	122	135	134
B	101	108	104
C	44	56	44
D_1	33	36	34
D_2	32	34	33
E	26	23	23
F	17.5	13.5	16.3
G	13.8	11.7	13.3
H	11.8	11.4	12.2
I	11.1	10.5	11.3
J	10.8	—	10.7

Based on staining intensities (for example, with Procion brilliant blue³⁰), components are present once per monomer with the exception of J, which in *Sulfolobus* and *Thermoproteus* seems to be present in four copies each.

Table 3 Total amounts, T_M s and ΔT_M s of 23S, 16S and 5S rRNA from *Thermoplasma* hybridized to archaeobacteria DNAs

DNA from	ng per 100 μ g DNA			rRNA			ΔT_M		
	23S	16S	5S	23S	16S	5S	23S	16S	5S
<i>Thermoplasma</i>	11.142	11.374	1.867	82.0	82.8	76.6	0	0	0
<i>Sulfolobus</i>	0.510	1.313	0.045	67.8	67.4	—	14.2	15.8	—
<i>Thermoproteus</i>	0.305	0.593	0.034	69.5	68.7	—	12.5	14.1	—
<i>E. coli</i>	0.014	0.031	0.018	—	—	—	—	—	—

³²P-labelled 23S, 16S and 5S rRNAs from *Thermoplasma acidophilum* were hybridized in $5 \times$ SSC and 25% formamide at 60 °C in excess over denatured DNAs from *Thermoplasma*, *Sulfolobus*, *Thermoproteus* and *E. coli* bound to nitrocellulose filters ($\sim 10 \mu$ g each per filter).

RNA polymerases⁶. This is in line with the presence of phytanol and C_{40} polyisoprenoid dialcohols in the lipids, the absence of murein in the wall of the organism and the failure of the antibiotics streptomycin, vancomycin and chloramphenicol to inhibit the growth of *Thermoproteus*²¹.

The striking homology of component patterns of the RNA polymerases of *Sulfolobus*, *Thermoplasma* and *Thermoproteus* (see Fig. 1, Table 2) is recognizable not only from the spacing but also from the staining properties of the components. The characteristic difference of these patterns from those of the RNA polymerases of *Halobacteria* and *Methanobacteria*, representing another type⁶, is a strong indication that *Thermoproteus*, like *Sulfolobus* and *Thermoplasma*, belongs to the thermoacidophilic rather than the methanogenic branch of archaeobacteria. This division is not immediately evident from the comparison of sequences of 16S rRNAs; for this criterion, *Sulfolobus* and *Thermoplasma* seem to be at least as distant from each other as from the methanogens.

The relationship of the three thermoacidophiles to one another is not close, as indicated by the lack of immunochemical cross-reaction of any one polymerase with antibodies against RNA polymerases of the other two in immunodiffusion experiments²⁴ (not shown). Similarly, no significant cross-reaction of RNA polymerases from Gram-negative or Gram-positive eubacteria has been observed by the same technique (K. O. Stetter, personal communication).

In addition to comparison of the partial sequences of 16S rRNAs, the relationships between different organisms have also been measured by heterologous hybridization^{25,26}. ³²P-labelled 23S, 16S and 5S ribosomal RNAs from *Thermoplasma* were hybridized to denatured DNAs from *Thermoplasma*, *Sulfolobus*, *Thermoproteus* and *E. coli* bound to nitrocellulose filters (see Table 3). The cross-hybridization of the large ribosomal RNAs with the DNAs from *Sulfolobus* and *Thermoproteus* was quite significant, although it was only 5–10% of the yield of the homologous hybridization. With both 23S and 16S rRNA the value for *Sulfolobus* DNA was somewhat higher than that for *Thermoproteus* DNA. No significant heterologous hybridization was observed between *Thermoplasma* 5S RNA and *Sulfolobus* or *Thermoproteus* DNA, or between all three rRNAs from *Thermoplasma* and *E. coli* DNA. Although the relative amounts of hybridized *Thermoplasma* rRNA suggest a slightly higher degree of homology of *Thermoplasma* to *Sulfolobus* than to *Thermoproteus*, the ΔT_M value in contrast shows a slightly higher residual homology of *Thermoplasma* to *Thermoproteus* than to *Sulfolobus*. As the association coefficient (S_{AH}) between *Thermoplasma* and *Sulfolobus* is about 0.17 (ref. 2), that of *Thermoplasma* and *Thermoproteus* can be expected to be of a similar size.

The homology of component patterns of the three thermoacidophiles to one another and the difference of these from those of members of the methanogens and halophiles, which also seem to be related to each other⁶, support the notion of two main branches of archaeobacteria, which has so far been based only on physiological and ecological evidence. The striking similarity of the component pattern of yeast RNA polymerases, especially polymerase A, to those of the thermoacidophiles suggests that the archaeobacteria, and particularly the ther-

moacidophiles, are more closely related to the eukaryotic cytoplasm than are the eubacteria. Archaeobacteria also possess other 'eukaryotic features'; these include the surprising sequence homology between initiator tRNAs of *Sulfolobus* and yeast (S. Nishimura, personal communication), the absence of formylation of methionyl initiator tRNA in archaeobacteria⁹, the ADP ribosylation of elongation factors (EF2s) from archaeobacteria by *Diphtheria* toxin¹⁰, the inhibition of both archaeobacteria and eukaryotes but not eubacteria by anisomycin²⁷ and the sequence homology between ribosomal A proteins of eukaryotes and archaeobacteria⁷. However, archaeobacteria do possess some eubacterial features, for example, the occurrence of restriction enzymes in *Thermoplasma* and *Sulfolobus* (ref. 28 and P. McWilliams, personal communication) and the existence of an RNA recognition sequence in 16S rRNAs of archaeobacteria²⁹.

The pattern of lipid components indicates some relationship between *Thermoproteus* and *Thermoplasma*²¹. The RNA polymerase component patterns, the lack of immunochemical cross-reaction between the RNA polymerases of all three thermoacidophiles and the similar distance of *Sulfolobus* and *Thermoproteus* from *Thermoplasma* support the conclusion that *Thermoproteus* represents a third order in the thermoacidophilic branch of archaeobacteria.

We thank Simon Wunderl for technical assistance and the Deutsche Forschungsgemeinschaft, SFB 51, for support.

Received 7 May; accepted 9 July 1981.

- Woese, C. R., Magrum, L. J. & Fox, G. E. *J. molec. Evol.* **11**, 245–252 (1978).
- Fox, G. E. *et al. Science* **209**, 457–463 (1980).
- Woese, C. R. & Gupta, R. *Nature* **289**, 95–96 (1981).
- Kandler, O. *Naturwissenschaften* **66**, 95–105 (1979).
- De Rosa, M., Gambacorta, A. & Bu'Lock, J. D. *Phytochemistry* **15**, 143–145 (1976).
- Zillig, W. & Stetter, K. O. in *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes* (eds Osawa, S., Ozeki, H., Uchida, H. & Yura, T.) 525–538 (University of Tokyo Press; Elsevier, Amsterdam, 1980).
- Matheson, A. T., Nazar, R. N., Willick, G. E., & Yaguchi, M. in *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes* (eds Osawa, S., Ozeki, H., Uchida, H., & Yura, T.) 625–637 (University of Tokyo Press; Elsevier, Amsterdam, 1980).
- Matheson, A. T., Moller, W., Amos, R. & Yaguchi, M. in *Ribosomes: Structure, Function and Genetics* (eds Chamblis, G., *et al.*) 297–332 (University Park Press, Baltimore, 1980).
- Pecher, T. & Böck, A. *FEMS Lett.* (in the press).
- Kessel, M. & Klink, F. *Nature* **287**, 250–251 (1980).
- Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R. & Wolfe, R. S. *Microbiol. Rev.* **43**, 260–296 (1979).
- Fox, G. E., Pechman, K. J. & Woese, C. R. *Int. J. syst. Bact.* **27**, 44–57 (1977).
- Magrum, L. J., Luehrs, K. R. & Woese, C. R. *J. molec. Evol.* **11**, 1–8 (1978).
- Brock, T. D., Brock, K. M., Belley, R. T. & Weiss, R. L. *Archs Microbiol.* **84**, 54–68 (1972).
- Darland, G. T., Brock, T. D., Samsonoff, W. & Conti, S. F. *Science* **170**, 1416–1418 (1970).
- Zillig, W., Stetter, K. O. & Tobien, M. *Eur. J. Biochem.* **91**, 193–199 (1978).
- Stetter, K. O., Winter, J. & Hartlieb, R. *Zentbl. Bakt. ParasitKde I* **201**, 214 (1980).
- Zillig, W., Stetter, K. O. & Janekovic, D. *Eur. J. Biochem.* **96**, 597–604 (1979).
- Zillig, W. *et al. Archs Microbiol.* **125**, 259–269 (1980).
- Sturm, S., Schönefeld, V., Zillig, W., Janekovic, D. & Stetter, K. O. *Zentbl. Bakt. ParasitKde I* **22**, 25 (1980).
- Zillig, W. *et al. Zentbl. Bakt. ParasitKde I* (in the press).
- Atanki, A. & Treter, R. *Am. J. clin. Nutr.* **25**, 1329–1332 (1972).
- Alberts, H., Amadio, J., Jenkins, M., Gutmann, E. & Forris, F. *Cold Spring Harb. Symp. quant. Biol.* **33**, 289–305 (1968).
- Ouchterlony, O. in *Frog. Allergy* **6**, 30–154 (1962).
- Ley, J. de & Smedt, U. de *Antonie van Leeuwenhoek* **287**, 307 (1975).
- Meyer, S. A. & Schlefer, K. H. *Archs Microbiol.* **117**, 183–188 (1978).
- White, T. J. & Bayley, S. T. *Biochim. biophys. Acta* **272**, 583–587 (1972).
- McConnell, D. J., Seay, D. G. & Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 1729–1730 (1978).
- Steitz, J. A. *Nature* **273**, 10 (1978).
- Fazekas, D. E., Groth, S. T., Webster, R. G. & Datynow, A. *Biochim. biophys. Acta* **71**, 377–391 (1963).
- Zillig, W., Zechel, K. & Halbwachs, H. *J. Hoppe-Seylers Z. physiol. Chem.* **381**, 221–224 (1970).
- Stembach, H., Engelhardt, R. & Lezius, A. G. *Eur. J. Biochem.* **60**, 51–55 (1975).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

BOOK REVIEWS

Three Mile Island re-visited

Alan Cottrell

IN APRIL 1980 the New York Academy of Sciences, in its highly commendable role of providing an open forum for the discussion of scientific questions, held a conference on the nuclear accident at Three Mile Island and the lessons to be learned from it. The results have been brought together in this book, Vol.365 of the *Annals of the New York Academy of Sciences*, which reports the conference.

The meeting was opened by a keynote address from Alvin Weinberg. To him, the main lesson of TMI was that "the central problem of nuclear energy is not waste disposal or even the connection between proliferation and nuclear power. It is, rather, reactor safety". This gave him the text for an admirably wise and perceptive analysis of the main problems in making nuclear power not only safe, but also acceptable to the general public. Writing about the situation in the USA (which he contrasted strongly with that in the UK and Europe generally) he argued eloquently and persuasively: first, for consolidating the present highly fragmented US power utility industry into fewer organizations, each large enough to handle the far-reaching responsibilities of reactor ownership and operation; and second, for putting power reactors all together in a small number of large and isolated sites, where they could be served by local communities who would be familiar, through their daily work, with nuclear energy. Weinberg thus called for a moratorium, not on new reactors, but on new reactor sites.

Many of the people who had been directly involved in the accident came to the conference and were able to give their personal views, as seen one year afterwards. Other participants, among whom were such familiar names as Norman Rasmussen and Chauncey Starr, included representatives of government and local authorities, the nuclear industry, the Nuclear Regulatory Commission (NRC), the President's Commission, universities and professional institutes, the media, public interest groups and residents of communities near Three Mile Island. They discussed many things: the technical history of the accident and the predicted likelihood of such an event; the amounts of radiation released and the problems, still to be solved, of decontaminating the site and removing safely the radioactive gases and water left in the buildings; the behaviour of the various individuals, organizations and authorities involved; the attitudes of the local inhabitants, some of whom remained

The Three Mile Island Nuclear Accident: Lessons and Implications. Edited by Thomas H. Moss and David L. Sills. Pp.343. ISBN pbk 0-89766-116-8. (New York Academy of Sciences: 1981.) \$66.

calm throughout, almost indifferent, whereas others suffered great mental distress.

A particularly lively part of the book deals with the acrimonious relations that have developed between the nuclear and public authorities and the press; the one side being accused of lying and concealing publically important information, the other of sensationalism, scare-mongering and ethical lapses. Whatever else TMI may have taught people, it has certainly exposed the atrocious state of communications between the nuclear industry and the general public via the media.

TMI was a Janus-faced accident, which told people what they wanted to know. To anti-nuclear people it showed how near the brink of public disaster the industry operates. To pro-nuclear ones it confirmed

the safety of the system, which demonstrably protected the public from radiation injury and all physical hazard.

Everyone — myself included — learned their own lesson from TMI. Mine was that, when something goes wrong with the cooling system of a pressurized water reactor, the temperature and pressure can then change so suddenly as to impose an almost impossible burden of instant reaction on the reactor operators. The serious questions which this raises seem to me to have been dodged, perhaps because they may even raise doubts about the basic concept of high-pressure water reactors. Only Weinberg dares in this book to ask the ultimate question: whether one should reconsider other reactor systems — such as gas-cooled reactors — that depend less on engineered safety features and more on inherent characteristics to prevent accidents. □

Sir Alan Cottrell, now Master of Jesus College, Cambridge, was previously the Chief Scientific Adviser to HM Government. His book, *How Safe is Nuclear Energy?*, was published by Heinemann Educational in June.

Digital image processing moves on

P.W. Hawkes

Two-Dimensional Digital Signal Processing. Edited by T.S. Huang. Part 1 *Linear Filters*, pp.230, ISBN 3-540-10348-1; Part 2 *Transforms and Median Filters*, pp.230, ISBN 3-540-10359-7. (Springer-Verlag: 1981.) Each Part DM 79, \$46.70.

ALTHOUGH there is no reason why development in the sciences should fit neatly into decades, it is a fact that the 1970s, the first decade of digital image processing, were dominated by a certain type of thinking whereas a rather different approach seems likely to pervade the coming years. The beginnings of the subject were strongly influenced by the algebraic work of Andrews, Hunt, Pratt and others associated with the University of Southern California, by the statistical work of Frieden and by Rosenfeld and his school at the University of Maryland. The first two tendencies were well-documented in an earlier book edited by Huang, *Picture Processing and Digital Filtering* (for review see *Nature* 264, 142), which the present volumes complement. The two parts are

very different: the first is devoted to two-dimensional filters, both non-recursive (R.M. Mersereau) and recursive (two chapters, by P.A. Ramamoorthy and L.T. Bruton and by B.T. O'Connor and T.S. Huang) and to two-dimensional Kalman filtering (J.W. Woods). The second is concerned with methods of performing operations frequently needed in image processing — matrix transposition (J.O. Eklundh), convolution and the discrete Fourier transform (H.J. Nussbaumer), Winograd's algorithm (S. Zohar) — and also with median filtering, statistical (B.I. Justusson) and deterministic (S.G. Tyan).

The break with the work collected in Huang's earlier volume is most clearly seen in Part I. With the exception of the chapter on recursive filtering, all of that earlier book could be read without conceptual difficulty and with great profit by anyone concerned with images, whether his background was physics, applied mathematics or electrical engineering; the chapter on recursive filtering was distinctly less accessible, partly because the authors did not provide an introduction addressed

to a wide audience but mainly because the topic was so very new. It is unfortunate therefore that in Part I of the new volumes, the authors of the chapter on the design of two-dimensional recursive filters assume that their audience is drawn wholly from the field of electrical engineering. That some of the jargon may not be familiar outside that field matters little; more serious is the fact that the authors do not attempt to relate their work to the types of situation where two-dimensional signals arise — images in particular. In the second contribution on recursive two-dimensional filters, O'Connor and Huang remedy this somewhat by giving considerably more introductory material, but a substantial effort is still required by the reader wishing to apply a recursive filter to an actual two-dimensional array. The same is true of the final chapter, on Kalman filtering, where once again, an introduction couched in slightly more general language would have rendered the material much easier to assimilate.

I have set out this criticism in some detail in view of the extremely interesting nature of the contents of these chapters. They contain much that has not hitherto been collected together and form a convenient and scholarly summary of recent developments in this field. If the publishers produce a revised paperback version of these volumes, as they did of Huang's earlier text, the addition of not-too-specialized introductory material to Part I at least would make it accessible to a much wider audience.

This criticism does not apply to most of the contributions to Part II, which deals mainly with efficient numerical implementation of mathematical operations. The first two topics, matrix transposition and fast calculation of convolutions and Fourier transforms, are such fast-moving fields that essays on them can never be perfectly up to date but it is nonetheless very useful to have these surveys of the subjects. Nussbaumer in his chapter on the use of polynomial algebra gives a readable account largely restricted to his own work — the chapter would have been more valuable if he had reviewed the other work on number-theoretic transforms, startlingly little of which appears in the bibliography but, such as it is, the text is clear and contains many examples. Zohar's chapter on Winograd's algorithm is likewise readable and well illustrated.

The final two contributions, on median filters, do not suffer from the hermetic tone of much of Part I. The purpose and use of such filters is made very clear and again, numerous examples of their use are provided.

All in all, these are important and interesting volumes, of the same high quality as their predecessor, though different in spirit. □

P.W. Hawkes is Maître de Recherches at the Laboratoire d'Optique Electronique, Toulouse.

Primates home on the Malayan range

Michael Kavanagh

Malayan Forest Primates: Ten Year's Study in Tropical Rain Forest. Edited by D. J. Chivers. Pp. 364. ISBN 0-306-40626-8. (Plenum: 1981.) £26.78, \$42.50.

DOUBTLESS most field primatologists who have originally set out to collect data for doctoral theses have seen other studies that needed doing in the same habitat, other behavioural and ecological questions that needed to be answered. But David Chivers is the first to have had the persistence to remain involved in the task of promoting such work for more than a decade and to have combined the major results of the first ten years of work into a single volume. The result is a unique book that describes the current state of knowledge of the monkeys and apes that are the dominant herbivores of South-east Asia's rain forests.

The book is intelligently organized, with the broader aspects of the scene being set first and followed by increasing levels of detail. Chivers's introduction briefly covers the main climatic and biological features of South-east Asia and traces the growth of primate studies in the area. Raemaekers, Aldrich-Blake and Payne then give an excellent overview of those features of the forest that are most salient to primate studies.

Three chapters then follow which, intentionally or not, illustrate the relative paucity of information on the socioecology of Malaysian monkeys compared with that on the lesser apes: Gittins and Raemaekers on gibbons, Curtin on leaf monkeys and Aldrich-Blake on macaques. Regrettably, terminology relating to range use shows some signs of confusion: gibbon "home ranges" are defined as the total area used in a given period, yet agile gibbons are mapped as wandering out of this three times in five days; and leaf monkey "territories" are defined as being the non-overlapping parts of home ranges shortly after a description of territorial overlap is given. Hunt's discussions of spacing mechanisms and determinants of range use also lack a certain clarity, especially in contrast with the Mackinnons' outstanding chapter on niche differentiation among the primates: here, observations and interpretations are clearly divided, areas of ecological separation and overlap are lucidly described and possible evolutionary pathways to the present community are discussed.

In the following contributions Fleagle relates locomotor and postural behaviour to ecology, Payne provides much useful data on non-primate competitors, and Chivers and Raemaekers give a unique description of long-term changes in the main study community, although they can be criticized for the lack of a clear division between speculation and certainty over the identity of some individual siamang. The

final chapter is an eclectic review that ranges from ecological parameters to broad conservation issues.

More detailed descriptions of methodology would have been helpful, but *Malayan Forest Primates* would make an excellent focus for advanced teaching. All quibbles aside, this is an essential reference work for primatology and tropical forest biology, containing much that is of importance for related disciplines. □

Michael Kavanagh is an Honorary Research Associate at the Veterinary Faculty, Universiti Pertanian Malaysia. He is currently a co-ordinator of the Malaysian Primates Research Programme.

Comprehensive cover

A.R. Archibald

Microbial Cell Walls and Membranes. By H.J. Rogers, H.R. Perkins and J.B. Ward. Pp. 564. ISBN 0-412-12030-5. (Chapman & Hall: 1980.) £30, \$75.

OLDER ideas that the surface layers of microorganisms were essentially inert, serving only to protect the interesting internal machinery of the cell, have long since given way to the realization that they are dynamic and plastic structures which are intimately involved in the functioning of the cell and its interaction with its external environment. Despite the increasingly widespread recognition of the importance of these surface layers, there has been a lack of an up-to-date and comprehensive text in which the topic is covered in some depth but at a level suitable for advanced undergraduate microbiologists and scientists in related areas.

The authors of *Microbial Cell Walls and Membranes* have remedied this omission and their book will be of interest to such readers as well as to scientists working in this field. Postgraduate students and others about to engage in studies on the structure and biosynthesis of cell walls, or on the mode of action of antibiotics that affect wall synthesis, will find the book particularly valuable. It provides a good introduction to the area, and gives a detailed account of the biochemistry of cell walls and a survey of membrane composition, structure and function. As the authors acknowledge, the various topics are discussed at somewhat different levels. The sections dealing with the structure and biosynthesis of bacterial cell wall components, the action of antibiotics and bacterial autolysins are particularly authoritative and well written.

The first two chapters provide a clear and well-illustrated account of the ultrastructure of bacterial walls and

membranes, and a useful description of procedures for their isolation and fractionation together with explanations of how these affect the nature and properties of the isolated material. The composition and functions of bacterial membranes are then discussed in some detail. The main emphasis, however, is on cell walls, particularly those of bacteria, though two fairly short chapters are included on the structure and biosynthesis of components of walls of yeasts and filamentous fungi.

Detailed attention is given to the structure and biosynthesis of the component polymers of walls. The interrelation and function of these components are discussed but some of the physical properties of walls and the surface properties of cells are less fully described. For example, the chapter on ultrastructure gives values for the thickness of the walls of several Gram positive bacteria but there is little reference to the evidence suggesting that the native hydrated walls are considerably thicker than those seen in fixed thin section. Brief reference is made to the ability of peptidoglycan to change in volume with ionic conditions but the effects of the presence of anionic polymers in the wall are not discussed, nor are the surface-charge properties of micro-organisms. No mention is made, either, of the flocculation of yeast, useful though that property has been over the last few thousand years.

While a fuller coverage of surface properties would require some consideration of capsules and surface appendages, I think that it would not have been out of place in a book such as this and could have been of interest and value, particularly to the undergraduate or more general reader. These might also have benefited from a rather more explicit description of some of the background material. Thus the final chapter, dealing with the assembly and growth of bacterial walls, makes only very brief reference to present or past thinking on the relation between genome segregation and wall growth and cell division: it does not describe topics such as the surface layers in growth and cell division in stalked and budding bacteria which, though not yet well understood at the molecular level, might have helped the non-expert reader to gain a wider perspective.

Opinions may differ on what would constitute an ideal balance but there is no doubt that this is a most useful and valuable book. The authors have faced a formidable task in selecting from the vast amount of information that is now available and have produced a clearly written and beautifully presented book, containing a wealth of references to reviews and original papers. It should interest a wide range of readers. □

A.R. Archibald is Professor of Microbiological Chemistry at the University of Newcastle upon Tyne.

Problems and solutions for phytoplankton

Karl Banse

The Physiological Ecology of Phytoplankton. Studies in Ecology, Vol.7. Edited by I. Morris. Pp.625. ISBN 0-632-00395-2/0-520-04308-1. (Blackwell Scientific/University of California Press: 1981.) £36, \$60.

MANY of us who work with phytoplankton indulge in speaking about "algae", "diatoms" or "phytoplankton", and in treating nutrient uptake or population growth in terms of the entire assemblage. This approach is in part forced upon us by our methods, for example the measurement of photosynthesis in a bottle of natural water. In part, however, we may be seduced by the apparent simplicity of a system of cells bathed individually by the environment and having high growth rates; they should be nearly in equilibrium with their surroundings. Investigators of larger, longer-lived invertebrates would hardly think of their objects of study in this way.

The last two chapters of the present volume, about succession and evolution, have reminded me about the individual phytoplankton species, of their problems of living, their solutions to these problems and of our ignorance. Further, in the other more conventional chapters I have been struck by the emergence and recurrent mention of studies of phase relations and time lags of responses in the laboratory, of the role of frequency of disturbance, and by the growing recognition of the importance of these phenomena in the field: algal cells are not necessarily in equilibrium with their surroundings but may lag behind the environmental change by a day or more.

The red thread running through this treatise is the interface between algal physiology and environmental factors. The subject has been tackled by commissioning in-depth, well-documented reviews which strive for a hydrobiological rather than a restricted marine scope. The first two chapters treat basic features of algal cells, symbioses and diseases, and general methodology. The latter subject re-surfaces, often in great detail, in several of the subsequent seven contributions on photosynthesis, nutrient dynamics, trace metals (iron, manganese and copper) and vitamins. The chemistry chapters, strongly physiological in their orientation, are not uniformly organized but all review modern data of the distribution of the substances in natural waters. The last six discuss sinking and flotation, patchiness of distributions, the role of algal cell size, grazing, succession and evolution. Although these chapters were subsumed by the editor under the general heading of population dynamics, many concern themselves principally with effects on specific growth rates, as did the preceding chapters by their very nature.

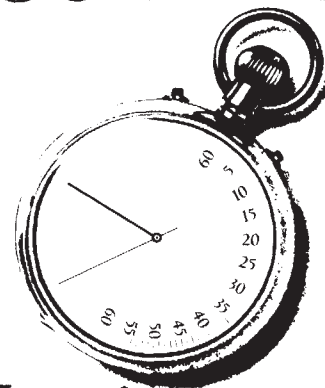
The book, however, is much more than a

collection of good to excellent but otherwise accidental reviews. The editor has achieved a broad coverage and, moreover, seems to have seen to it that there are appropriate cross-references between chapters. The large index is helpful but could have been better organized. For example, the subject of "biomass estimations" does not appear as such but as a subheading under the three species studied in this context, whereas the text referred to gives a general account of the subject. Likewise, "copper inhibition" is not to be found under "copper" — so even today one can overdo the species bit!

In my opinion, this volume will be useful to the phytoplankton field worker (marine and freshwater), the advanced, general ecologist and the physiologist interested in ecological aspects of the subject. It is likely to be read or consulted for quite some time even though almost all manuscripts were apparently written in 1977 and the authors had little opportunity to add later references. The chapters are largely so specialized that the book as a whole cannot be used as a text; however, some contributions are to be recommended as a special reading for more senior students. □

Karl Banse is Professor of Oceanography at the University of Washington, Seattle.

IRCS JOURNALS



Keeping pace with science

IRCS Medical Science
St. Leonard's House
St. Leonardgate,
Lancaster, U.K.

A member of Elsevier Science Publishers

The past is the key to the present?

Janet V. Watson

Precambrian of the Southern Hemisphere. Developments in Precambrian Geology, 2. Edited by D.R. Hunter. Pp.882. ISBN 0-444-41862-8. (Elsevier Scientific: 1981.) \$170.75, Dfl 350.

THE discovery, a decade or so ago, that the crustal regime that controls present-day geological processes operates on a world-wide scale gave earth scientists a much-needed push towards thinking of rock structures in terms of large, simple patterns. This approach worked well for the oceanic crust, where the scale of the magnetic striping matches that of the ocean basins themselves. But the continents are another matter. The time-span of continental evolution is more than ten times that recorded in oceanic crust and the continental crust is a patchwork of units formed during different periods in response to entirely different controls.

Thus every land-based geologist has at some time come up against the problem of how to make sense of a complex terrain in regional or global terms without doing violence to the evidence of his own eyes. The authors of this book have tackled the problem with remarkable success — they have at once produced a work of reference which will stand for many years to come and revealed new patterns and relationships on a continental scale. If it is true that a child's first seven years set the pattern for the rest of its life, it must surely be true that the fundamental features of the continental crust were built into it in Precambrian times. One of the fascinations of the book is that although Phanerozoic events such as the development of the African rift valley are not dealt with directly, they are foreshadowed in accounts of structures formed two billion years earlier.

The scene is set in an editorial introduction and in three admirable chapters introducing the sections on Australia, Africa south of the equator and South America. A welcome feature, seen especially in the contributions by R.W.R. Rutland and by D.R. Hunter and D.A. Pretorius, is the integration of geophysical and geological data to document the extent and boundary relations of successive tectonic provinces. These readable and thought-provoking chapters establish the large-scale structural patterns which must be explained in any global hypothesis of crustal evolution. The details are filled in by chapters which deal, area by area, with succession, structure, magmatism and mineralization. Sedimentary environments and palaeontology (often neglected by Precambrian geologists) take their place alongside the more traditional aspects. Although it would be idle to claim that all of these chapters make easy reading, there are abundant rewards. The long familiarity of the various authors with their areas and

their self-denial with respect to theoretical hobby-horses give authority to the text, which covers areas such as Zimbabwe that have profoundly influenced our understanding of the Precambrian crust. The juxtaposition of apparently unrelated observations gives unexpected insights into the connections between older and younger structures, or between magmatism and basin formation — I particularly liked A.D.T. Goode's suggestion that certain types of Archaean belt seem destined to remain unstable.

The supercontinent of Gondwanaland has been crucial to thoughts on continental drift for more than 60 years. Alex du Toit,

that wisest Gondwanaland geologist, gave a third of his *Geology of South Africa* to the Precambrian at a time when it was customary to dismiss pre-Phanerozoic events in a page or so. Now, Dr Hunter and his colleagues have built up the Precambrian picture into a coherent whole. The lion's share of the book rightly goes to Australia and Africa. India excluded itself long ago by drifting into the Northern Hemisphere, South America receives somewhat cursory treatment and Antarctica is, rather sadly, omitted. But the core of the Gondwanaland story is now there for all to read. Du Toit would surely have approved. □

Janet V. Watson is a Professor in the Department of Geology at Imperial College, University of London.

Mixed relations between yeasts and man

James A. Barnett

Biology and Activities of Yeasts. The Society for Applied Bacteriology Symposium Series, No. 9. Edited by F.A. Skinner *et al.* Pp.310. ISBN 0-12-648080-X. (Academic: 1981.) £15.80, \$38.

RIGHTLY, for a symposium of an "applied" society, half of the chapters in this book concern industrial problems. There are remarkable yeasts that can ferment honey of ~80% hexose, and Tilbury's careful review of such yeasts that tolerate high concentrations of sugar or salt is particularly welcome. Unfortunately, Davenport's survey of cold-tolerant yeasts is not comparable; such is still needed, even if only to summarize the physiological research required to solve current industrial problems. Writing on spoilage, Seiler (bakery products) and Dennis and Buhagiar (fruit and vegetables) bring out the anecdotal character of many "applied" studies: even when researchers trouble to identify the yeasts properly, often no physiological work on their role is undertaken. An exception is the last author's work on the breakdown of cellulose of rhubarb by *Trichosporon* species.

There are brief, helpful, summaries of work on the flocculation and ethanol tolerance of *Saccharomyces cerevisiae* (Rose). Put and de Jong review the heat-resistance of yeasts that spoil soft drinks and other products of fruits. In addition, laymen should be grateful for two authoritative medical articles: Hurley describes the increasingly recognized importance of infections by yeasts of the genus *Candida*, while Campbell and Mackenzie give an admirably concise up-to-date survey of *Cryptococcus neoformans*.

Three chapters are taxonomical, but more for taxonomists than for applied microbiologists. For example Kreger-van Rij (generic differences) complains,

reasonably enough, about abolishing *Torulopsis* and, elliptically, expresses disapproval of the new genus *Hormoascus*. According to von Arx, yeasts are deficient moulds and, if the wicked twentieth-century microbiologist, Kluyver, hadn't got his hands on them, yeasts would now be properly classified, according to their anatomical structures and the strict legal rules of Botanical Nomenclature. Both of these taxonomists, and also van der Walt, in his excellent discussion of delimiting species, believe in phylogenetic classification and are reverent towards such measurements as "%G+C", ignoring possible differences in rates of evolution of DNA base composition. These subjects, suitable for private intercourse between consenting taxonomists, were perversely selected for this symposium: taxonomic problems in applied work are chiefly those of identifying, not classifying, yeasts.

Furthermore, although the editors boast of this volume as a "comprehensive review of the biology and activities of yeasts", they omit the genetic (nuclear and mitochondrial) and physiological control of metabolism, as well as other current research of potential industrial interest, such as genetic manipulation by protoplast fusion and recombinant DNA techniques, and also C₁-metabolism.

Most of this symposium should be useful to teachers of microbiology, and specialists may be interested in certain articles, notably those on taxonomy and "xero-tolerant" yeasts. Accordingly, the book should be in university and technical college libraries, as well as in those serving institutes for applied research where yeasts are studied. □

James A. Barnett, employed by the Agricultural Research Council, is honorary lecturer in the School of Biological Sciences at the University of East Anglia.